

A;Molecule type: DNA
A;Residues: 1-258 <WIL>
A;Cross-references: UNIPARC:UPI0000164318; EMBL:Z78417; PIDD: CAB01692.1; GSEDB:

```

C:GeneID: CESP:C35C5.3b
A:Map position: X
A:Introns: 11/3; 34/1; 67/2; 103/2; 144/3

Query Match          15.4%; Score 92; DB 2; Length 258;
Best Local Similarity 18.0%; Pred. No. 0.21;
Matches 24; Conservative 18; Mismatches 23; Indels 68; Gaps 2;

Oy 29 YCBPGFEATYLELASAVEQYPGIEIESR-----LGG- 61
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 102 YCVSCGYKQADPDQFTFAKEKYPNNPIEGANFAPVLMKAYVAQALFVKMAYLVLVGGI 161
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy 62 -----TGAPEIINQVLPFSKLEN 80
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 162 NPFERFGLGYQIILQAHGNKNSCMLVMTLGNLVESQLISTGAPEYVLGNQIWSKIES 221
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy 81 GGFPEYKDLIEAI 93
      |||:::|
Db 222 GRVPSPOEFMQLI 234

RESULT 3
AD2604
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AD2604
R:Wood, D.W.; Seubald, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCelligott,
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2604
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-101 <KOR>
A:Cross-references: UNIPROT:Q8U9R5; UNIPARC:UPI00000D173B; GB:AE008688; PIDD:AA141250.1;
A:Experimental source: strain C58 (Dupont)
C:GeneticB:
A:Gene: Atu0228
A:Map position: circular chromosome

Query Match          15.1%; Score 90; DB 2; Length 101;
Best Local Similarity 29.6%; Pred. No. 0.11; 34; Indels 6; Gaps 3;
Matches 24; Conservative 17; Mismatches 34; Indels 6; Gaps 3;

Oy 16 EVDGSGRVIVYECBPGFEATYLELASAVEQYPG-IEIESRLGGTGA-FEIEINGOL 73
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3 ETKR-----RIAIRCTQCNMLLRAGWMAQELIQFASDIGVSLPISGGLFETVDTGI 58
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|

Oy 74 VFSKLENGGFPYEKDLIEAIR 94
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 59 IWERKRDGGFPPEKELKORIR 79

RESULT 4
D97386
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: D97386
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D97386
A:Status: preliminary
A:Molecule type: DNA

```

A;Residues: 1-101 <KTR>
A;Cross-references: UNIPROT:Q8UIR5, UNIPARC:UPI00000D173B, GB:AEO07869, PIDN:AAK86045.1,
C|Genetics:
A;Gene: AGR_C_387
A;Map position: circular chromosome

Query Match 15.1%; Score 90; DB 2; Length 101;
Best Local Similarity 29.6%; Pred. No. 0.11;
Matches 24; Conservative 17; Mismatches 34; Indels 6; Gaps 3;

OY 16 EVEDSGVRIVVEYCEPGFEATYLELASAVKEQYPG-IEISRLGGTGA-FEIEINGOL 73
 |||:||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3 ETKP----RAIRYCCTGCWMLLRAGMWAQEILLQTFSADIGEVSLIPISTGLFETIVDGTI 58
 :::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

DY 74 VFKTLNGGFPEEKDLLEAIR 94
 :::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 59 IWERKRDDGFPPKKELRKORV 79

RESULT 5
DB3228
hypothetical protein PAJ338 [imported] - Pseudomonas aeruginosa (strain PAO1)
C|Species: Pseudomonas aeruginosa
C|Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C|Accession: D83228
R|Stover, C.K.; Yam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Breyer, S.; Yun, Y.I.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; Lim,
Adman, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
A;Reference number: AB2950; PMID:20437337; PMID:10984043
A;Accession: D83228
A;Status: preliminary
A|Molecule type: DNA
A;Residues: 1-96 <STD>
A;Cross-references: UNIPROT:Q9HYQ7, UNIPARC:UPI00000C59C0, GB:AEO04756, GB:AEO04091, NIDN:
A;Experimental source: strain PAOI
C|Genetics:
A;Gene: PAJ338

Query Match 13.9%; Score 83; DB 2; Length 96;
Best Local Similarity 31.2%; Pred. No. 0.49;
Matches 25; Conservative 13; Mismatches 36; Indels 6; Gaps 3;

DY 19 PGSGRVIVEYCPGE--EATYL--ELASAVEKYPTGISERLGGTGAFEEINGQLV 74
 |||:||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2 PLAKPFIIVTYCTQCWWLLFAAWLAQELLSTPADLDGKVCLBPGTG--GVFRITCDGVGV 59
 :::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 75 FSKLNGGFPEEKDLLEAIR 94
 :::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 60 WERRKDGFPEEARALKORV 79

RESULT 6
DB3257
selenoprotein W-related protein VC0982 [imported] - Vibrio cholerae (strain N16961 seroogy
C|Species: Vibrio cholerae
C|Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C|Accession: J82257
R|Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Peterson,
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, P.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: AB2035; PMID:20406833; PMID:10952301
A;Accession: J82257
A>Status: preliminary
A|Molecule type: DNA
A;Residues: 1-97 <HEI>
A;Cross-references: UNIPROT:Q9KTCl, UNIPARC:UPI00000C2E43, GB:AEO04179, GB:AEO03852, NIDN:
A;Experimental source: seio group O1, strain N16961, biotype El Tor
C|Genetics:
A;Gene: VC0982

C:Genetics:
A:Gene: EMBL10262
A:Map position: 11

Query Match 13.1%; Score 78.5; DB 2; Length 101;
Best Local Similarity 26.9%; Pred. No. 1.4;
Matches 21; Conservative 16; Mismatches 34; Indels 7; Gaps 2;

QY 24 RIVVEYCEPGFATYLELASAVKEQYPC--TEISRLGGTGAFEIEI-----NGQLVFS 76
DB 6 RISTYTCQCQMLLRAAMAGELLOTFGODLAEVALRPETGVFEIRVMPDSESLIVE 65

QY 77 KLENGGFPYKDLIEAIR 94
DB 66 RKDDGFPPEAKYVKORVR 83

RESULT 11

T36249

CDA peptide synthetase II SCE63.02c [imported] - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 12-Jul-2004

C:Accession: T36249
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999

A:Reference number: 221602

A:Accession: T36249

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3670 <SAU>

A:Cross-references: UNIPROT:Q924X5; UNIPARC:UP100000DAFLE; EMBL:AL035640; PIDN:CA838517.

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: cds82, SCOEDB:SC63.02c

C:Superfamily: non-ribosomal peptide synthetase; acetate-CoA ligase homology; acyl carrier

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:512-932/Domain: acetate-CoA ligase homology <AC11>

F:948-1061/Domain: acyl carrier protein homology <ACP1>

F:1545-1981/Domain: acetate-CoA ligase homology <ACP2>

F:1997-2065/Domain: acyl carrier protein homology <ACP2>

F:2608-3064/Domain: acetate-CoA ligase homology <AC13>

F:3080-3147/Domain: acyl carrier protein homology <ACP3>

F:980,2029,3112/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 13.1%; Score 78; DB 2; Length 3670;
Best Local Similarity 31.7%; Pred. No. 97;
Matches 26; Conservative 14; Mismatches 30; Indels 12; Gaps 3;

QY 39 YLELASAVKEQYPGIEISRLGGTGAFEIEINGQLVFSKLENGGFPYKDLIEAIRASN 98
DB 2740 YLAVSA---ELYPGAVGNALHSPISFDLTVTG--LFPALNCGCVHLADLEHLARALD 2794

QY 99 GE-----TLEKITSRPPCV 113
DB 2795 GEVPLDPTTFKATPSHPLI 2816

RESULT 12

T21526
hypochemical protein F28H7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T21526
R:Berks, M.
submitted to the EMBL Data Library, May 1996
A:Reference number: 219435
A:Accession: T21526
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-232 <WIL>
A:Cross-references: UNIPROT:Q19892; UNIPARC:UP10000335A34; EMBL:Z72508; PIDN:CAA96637.1;
A:Experimental source: clone F28H7
C:Genetics:

A:Gene: CESP:F28H7.4
A:Map position: 5
A:Introns: 23/1; 134/3

Query Match 12.7%; Score 76; DB 2; Length 232;
Best Local Similarity 15.1%; Pred. No. 6.4;
Matches 21; Conservative 19; Mismatches 31; Indels 68; Gaps 1;

QY 23 RIVVEYCEPGFATYLELASAVKEQYPCIEI----- 55
DB 86 INLTVSVSCGYQAFQAFYEFPAKEXKPGLVIGGNFSPDFWKGCLAQIVGAKIGILA 145

QY 56 -----ESRLGGTGAFEIEINGQLV 74
DB 146 IVITGSPNFEYIGFGPQIQLTAHYNRFSYSLVFMIGNLFESTLSSTGAFEIFGDKOI 205

QY 75 FSKLENGGFPYKDLIEAI 93
DB 206 WSKISKERVPTQEEFLNLI 224

RESULT 13

B64485
phosphoribosylglycinamide formyltransferase 2 (EC 2.1.2.-) M01486 [similarity] - Methanococcus

C:Species: Methanococcus jannaschii

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: B64485

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurec, M.A.

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A:Reference number: A63300; MUID:96337999; PMID:8688087

A:Accession: B64485

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-393 <BDL>

A:Cross-references: UNIPROT:Q58881; UNIPARC:UP10000332B6F; GB:U67589; GB:L77117; NID:928

A:Map position: REV1458908-1457727

A:Start codon: TTG

A:Function:

A:Description: catalyzes the production of beta-formyl glycinamide ribonucleotide from f

A:Note: cofactor magnesium

C:Superfamily: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain; f

C:Keywords: magnesium; purine nucleotide biosynthesis; transferase

F:29-371/Domain: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain

Query Match 12.6%; Score 75.5; DB 1; Length 393;
Best Local Similarity 22.5%; Pred. No. 13;
Matches 31; Conservative 23; Mismatches 37; Indels 47; Gaps 7;

QY 1 MSGPGQTSVAPPEEVEP-----GSGVRIVE-----Y 29
DB 161 MSSSGKQSVRSBEDLEKMKIKEGARGIGNRVIEFINPFIETLLTAPATGKTF 220

QY 30 CEPGC---FEATYLE-----LASAVKEQYPCG--TEISRLGGTGAFEIEI---NGQLVF 75
DB 221 CEPIGHVIDDDYHESQPHMSAELKEQODIAKVTDALGGYIGVVELFVKGDEVIF 280

QY 76 SKLENGGFPYKDLIEAI 93
DB 281 SEVSPR--PHOTGMVTMI 296

RESULT 14

AD2985
oxidoreductase Atu3485 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AD2985

A:Gene: At3g3485
A:Map position: linear chromosome
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577, MUID:21608550; PMID:11743193
A:Accession: AD2985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-698 <KUR>
A:Cross-references: UNIPROT:O8UHA9, UNIPARC:UPI0000016480F, GB:AE008689, PIDN:AAL44298.1
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: At3g3485
A:Map position: linear chromosome

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2006, 12:00:46 ; Search time 135.659 Seconds
(without alignments)
598.087 Million cell updates/sec

Title: US-09-824-787B-2
Perfect score: 597
Sequence: 1 MSGEPQTSVAPPEEVEPC.....ASNGETLEKITSRPPCVIL 115

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	115	2 Q9BRT3 HUMAN	Q9BRT3 homo sapien
2	537	89.9	115	2 Q9C086 MOUSE	Q9C086 m mus muscu
3	435.5	72.9	126	2 Q5ZRH7 CHICK	Q5ZRH7 gallus gall
4	273	45.7	95	2 Q80285 BRARE	Q80285 brachydanio
5	221.5	37.1	94	2 Q80288 BRARE	Q80288 brachydanio
6	218	36.5	89	2 Q4T504 TETNG	Q4T504 tetraodon n
7	148.5	24.9	95	2 Q9VRA0 DROME	Q9VRA0 drosophila
8	137.5	23.0	88	2 Q8H6T4 CHARE	Q8H6T4 chlamydomon
9	137	22.9	80	2 Q4UBK6 SULAC	Q4UBK6 sulfolobus
10	132.5	22.2	93	2 Q7QFL8 ANOCA	Q7QFL8 anopheles g
11	119	19.9	232	2 Q8S227 ORYGA	Q8S227 oryza sativ
12	114.5	19.2	244	1 HSP61 HERGL	Q8B237 oryza sativ
13	112.5	18.8	86	1 SEPM1 HUMAN	Q8B237 oryza sativ
14	112.5	18.8	86	1 SEPM1 HUMAN	Q8B237 oryza sativ
15	112.5	18.8	86	1 SEPM1 HUMAN	Q8B237 oryza sativ
16	112.5	18.8	86	2 Q802G9 BRARE	Q802G9 brachydanio
17	112.5	18.8	87	2 Q5NVB2 PONYA	Q5NVB2 pongo pygma
18	111	18.6	209	2 Q67ZS9 ARATH	Q67ZS9 arabidopsis
19	110.5	18.5	209	2 Q6N0E9 ARATH	Q6N0E9 arabidopsis
20	110.5	18.5	346	1 SELV HUMAN	Q59797 homo sapien
21	109	18.3	329	2 Q80T12 MOUSE	Q80T12 mus muscu
22	107.5	18.0	87	1 SEPM1 RAT	Q63301 mus muscu
23	107.5	18.0	257	1 Q6IS17 ORYGA	Q6IS17 oryza sativ
24	104	17.4	257	1 Q6IS17 ORYGA	Q6IS17 oryza sativ
25	103.5	17.3	228	2 Q8W1E5 ARATH	Q8W1E5 arabidopsis
26	99	16.6	92	2 Q8DFL8 VIBVU	Q8DFL8 vibrio vuln
27	99	16.6	94	2 Q7WMS1 VIBVU	Q7WMS1 vibrio vuln
28	98.5	16.5	86	1 SEPM1 SHEEP	Q19037 ovis aries
29	94	15.7	92	2 Q87RH7 VIBPA	Q87RH7 vibrio para
30	94	15.7	103	2 Q60CA7 METCA	Q60CA7 methylococc
31	93	15.6	93	2 Q92SUT7 RHIME	Q92SUT7 rhizobium m

32	93	15.6	201	2 Q7PVR5 ANOCA	Q7PVR5 anopheles g
33	92.5	15.5	163	2 Q802F2 BRARE	Q802F2 brachydanio
34	92	15.4	81	2 Q70ZL6 GIALA	Q70ZL6 giardia lam
35	92	15.4	247	1 SELT1 CAEBR	Q9JN55 caenorhabdi
36	92	15.4	247	1 SELT1 CAEBR	Q9JN55 caenorhabdi
37	90	15.1	90	2 Q5YEV1 CHLS6	Q5YEV1 chlorella
38	90	15.1	101	2 Q8VIR5 AGRTS	Q8VIR5 agrobacteri
39	89.5	15.0	1125	2 Q5U4C0 MOUSE	Q5U4C0 mus muscu
40	89.5	15.0	1137	2 Q8B159 MOUSE	Q8B159 mus muscu
41	89.5	15.0	1192	2 Q8B179 MOUSE	Q8B179 mus muscu
42	88	14.7	195	1 SELT HUMAN	Q62341 homo sapien
43	88	14.7	145	1 SELT MOUSE	Q62342 mus muscu
44	87.5	14.7	144	2 Q4RYD8 TETNG	Q4RYD8 tetraodon n
45	87	14.6	65	2 Q57WJ2_9TRYP	Q57WJ2 trypanosoma

ALIGNMENTS

RESULT 1
ID Q9BRT3_HUMAN PRELIMINARY; PRT; 115 AA.
AC Q9BRT3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)
DE Chromosome 17 open reading frame 37 (XTP4) (c35 protein).
GN Name=C1orf137; Synonym=XTP4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schliefer G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helen F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Millar S.J.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.C., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butcherfield J.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schnerker A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RA Director MGC Project;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RL Liu Y., Cheng J., Lu Y., Wang G., Zhang L., Chen J., Li L.,
Submitted (Mar-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Evans E.E., Heun A.D., Luhovsky S., Paris M.J., Borrello M.A.,
RA Smith E.S., Sahnasrabudhe D.M., Zauderer M.,
Submitted (Dec-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006006; AA06006.1; -, mRNA.
EMBL; AF490253; AA085461.1; -, mRNA.

DR EMBL; AY508814; AAR92035.1; -; mRNA.
 DR Ensemble; ENSG00000141741; Homo sapiens.
 DR HGNC; HGNC:28230; C17orf37.
 DR InterPro; IPR011893; CXXU_selWTH.
 DR TIGRFAMs; TIGR02174; CXXU_selWTH.
 DR SEQUENCE 115 AA; 12403 MW; 5DB8911C0F23DC1 CRC64;
 Query Match 100.0%; Score 597; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 7,5e-51;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSGEGQTSVAPPEEVEPGSGVRIIVEYCEPCGEATYELBLASVKEQYPCIEISRLG 60
 DB 1 MSGEGQTSVAPPEEVEPGSGVRIIVEYCEPCGEATYELBLASVKEQYPCIEISRLG 60
 QY 61 GTGAFEINQVSKENGEPPEKLEIRASNGETLEKTNRPPCVIL 115
 DB 61 GTGAFEINQVSKENGEPPEKLEIRASNGETLEKTNRPPCVIL 115
 RESULT 2
 Q9CQ86_MOUSE PRELIMINARY; PRT; 115 AA.
 AC Q9CQ86;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
 DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
 DE library, clone:2310051B14 product:hypothetical protein, full insert
 DE sequence (Mus musculus adult male brain cDNA, RIKEN full-length
 DE enriched library, clone:0710001H16 product:hypothetical protein, full
 DE insert sequence) (Mus musculus 10 day old male pancreas cDNA, RIKEN
 DE full-length enriched library, clone:1810046J19 product:hypothetical
 DE protein, full insert sequence) (Mus musculus adult male aorta and vein
 DE cDNA, RIKEN full-length enriched library, clone:A530099C24
 DE product:hypothetical protein, full insert sequence) (RIKEN CDNA
 DE 1810046J19).
 OS Mus musculus (Mouse).
 GN Name=1810046J19R1k;
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein, Brain, Pancreas, and Tongue;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency long-length cDNA cloning";
 RL Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein, Brain, Pancreas, and Tongue;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamazaki I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Straub P., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guelzim S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontecki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:695-690(2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein, Brain, Pancreas, and Tongue;
 RX MEDLINE=22354683; PubMed=12468851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nishida I., Oseko N., Saito R., Suzuki H., Yamazaki I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schoenbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schiraldi L.M., Knapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Brad D., Bruce V., Chochia C., Corbett L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Guclincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierzki R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sanderlin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 RA Hiraoka T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume M., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein, Brain, Pancreas, and Tongue;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein, Brain, Pancreas, and Tongue;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishogi K.,
 RA Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Oyeda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
 RA Kozeki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-34-format
 RT sequencing pipeline with 384 multicapillary sequencer";
 RL Genome Res. 10:11757-11771(2000).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Brain, Pancreas, and Tongue;
 RA STRAIN=C57BL/6J; TISSUE=Brain, Pancreas, and Tongue;
 RA Aachai P., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hasegawa T., Hara A., Hayatsu N., Hiramoto K., Hiraka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawaji H., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tjolina Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]

RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
 RA Aachai J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiroka T., Hirozane T.,
 RA Horii F., Imcanti K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai T., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ono M., Ohashi N., Ozaki Y.,
 RA Saito R., Saitoh H., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
 RN Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [8]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Liver;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derje J.G.,
 RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marinsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueidi T.B., Toshikiyuki S., Carninci P., Pirange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Sannerfeld J., Schein J.E., Jones S.J.W., Marra W.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [9]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=FVB/N; TISSUE=Liver;
 RA Strausberg R.;
 DR Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK009922; BAB26586.1; -; mRNA.
 DR EMBL: AK009859; BAB22480.1; -; mRNA.
 DR EMBL: AK007795; BAB25261.1; -; mRNA.
 DR EMBL: AK041314; BAC30901.1; -; mRNA.
 DR EMBL: BC021589; AAB21589.1; -; mRNA.
 DR EMBL: ENSMUSG0000002580; Mus musculus.
 DR MGI: MGI:1913678; 1810046319Rik.
 DR InterPro: IPR011893; CXXU_selWTH.
 DR TrEMBL: TIGR02174; CXXU_selWTH; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 115 AA; 12295 MW; B36A0340DFBA737A CRC64;
 Query Match 89.9%; Score 537; DB 2; Length 115;
 Best Local Similarity 89.6%; Pred. No. 6.1e-45; Indels 0; Gaps 0;
 Matches 103; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MSSEPGTSTVAPPEVEPSPGVRIVVEYCEPCGFATYIELASAVEQYPGIEISRLG 60
 DB 1 MSSEPPAVSVVPPPGVEAGSGVHIVVEYCKPCGFATYIELASAVEEYPGIEISRLG 60
 QY 61 GTGAFETIEINGQLVFSKLENGSPPEYKDLIRAIRASNGSTLEKITSRPPCVIL 115
 DB 61 GTGAFETIEINGQLVFSKLENGSPPEYKDLIRAIRASNGSPVEKITSRPPCVIL 115
 RESULT 3
 ID Q52IH7 CHICK
 AC Q52IH7/
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=RCU0804.26b2;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phaethonidae; Phaethonidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CB; TISSUE=Bursa;
 RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezubov Y., Zaim J.,
 RA Pledier P., Kutter S., Biagodatski A., Kostovska D., Kotter M.,
 RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.,
 RT "Full-length cDNAs from chicken bursa lymphocytes to facilitate
 RT gene function analysis";
 RL Genome Biol. 6:R6-R6(2005).
 DR EMBL: AJ220807; CAG32466.1; -; mRNA.
 DR InterPro: IPR011893; CXXU_selWTH.
 DR TrEMBL: TIGR02174; CXXU_selWTH; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 126 AA; 13438 MW; 12B8498FD40DBB6D CRC64;
 Query Match 72.9%; Score 435.5; DB 2; Length 126;
 Best Local Similarity 72.2%; Pred. No. 6.9e-35;
 Matches 91; Conservative 4; Mismatches 16; Indels 15; Gaps 2;
 QY 2 SGEPGCTSVAPPEVEPSPGSG-----VRIVEYCEPCGFATYIELASAVEKEQ 49
 DB 4 SGNGNGAAAG---TSEADGDGFGSDSGSERVHIVVEYCEPCGFATYIELASAVEE 60
 QY 50 YPGIEIESLGGTGAIEIINGQLVFSKLENGSPPEYKDLIRAIRASNGSTLEKITSR 109
 DB 61 YPGIEIESLGGTGAIEIINGQLVFSKLENGSPPEYKDLIRAIRASNGSTLEKITSR 120
 QY 110 PCCVIL 115
 DB 121 PCCVIL 126
 RESULT 4
 ID Q802F5 BRARE
 AC Q802F5/
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Selenoprotein W2a.
 GN Name=seppw2a;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC MEDLINE=22796661; PubMed=12915322; DOI=10.1016/S1567-133X(03)00054-1;
 RA Thilse C., Degraeve A., Kryukov G.V., Gladyshev V.N.,
 RA Obrecht-Pflumio S., Krol A., Thilse B., Lescure A.,
 RT "Spatial and temporal expression patterns of selenoprotein genes
 RT during embryogenesis in zebrafish";
 RL Gene Expr. Patterns 3:525-532(2003).
 DR EMBL: AY221261; AAO65270.1; -; mRNA.
 DR Ensemble: ENSDARG0000027296; Danio rerio.
 DR ZFIN: ZDB-GENE-030428-1; seppw2a.
 DR InterPro: IPR011893; CXXU_selWTH.
 DR TrEMBL: TIGR02174; CXXU_selWTH; 1.
 KW Selenium; Selenocysteine.
 FT SE_CYS 13
 SQ SEQUENCE 95 AA; 10572 MW; 05A25E769DDFDD0B CRC64;
 Query Match 45.7%; Score 273; DB 2; Length 95;


```

Query 22 GVRIVERCEPCGFATYLELSAVKEQYPGIEISRLGGTGAFFIEINGQLVFSKLENG 81
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 2 GQAIKVEYCGCGGCGEPYRQELKRVVTAFTDADVGFGVQSGFPIEINGQLVFSKLETS 61
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 82 GPVEKDLIEAIRASNGETLEKINSRPPCVIL 115
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 62 GPVEDDIMGVIGRAYDQGPVEKITKSGPPCVIL 95
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 5
0802G8_BRARE
ID 0802G8_BRARE PRELIMINARY; PRT; 94 AA.
AC 0802G8;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Selenoprotein W2b.
GN Name=sepw2b;
OS Brachydonto rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxID=7955;
RX NCBI_TaxID=7955;
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22796661; PubMed=12915322; DOI=10.1016/S1567-133X(03)00054-1;
RA Thisse C, Degraeve A, Kryukov G.V., Gladyshev V.N.,
RA Obrecht-Pflumbo S., Krol A., Thisse B., Lescaur A.;
RT "Spatial and temporal expression patterns of selenoprotein genes
during embryogenesis in zebrafish.";
RL Gene Expr. Patterns 3:525-532(2003).
DR EMBL; AY216583; AAC86697.1; -; mRNA.
DR Ensembl; ENSDARG0000014390; Danio rerio.
DR ZFIN; ZDB-GENE-030428-2; sepw2b.
DR InterPro; IPR011893; CXXU_selMTH.
DR TIGRFAMs; TIGR02174; CXXU_selMTH; 1.
KW Selenium; Selenocysteine.
FT SE_CYS 13
FT 13
SQ SEQUENCE 94 AA; 10622 MW; C92468CBF5E2655 CRC64;

Query Match 37.1%; Score 221.5; DB 2; Length 94;
Best Local Similarity 46.2%; Pred. No. 6e-14;
Matches 43; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 23 VRIVVECPGCGFEATYLELSAVKEQYPGIEISRLGGTGAFFIEINGQLVFSKLENG 82
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3 VKVIEIYCGAGYERFQELKREICNGCPDAVSGFVGRCGFQINDLVPFSKLESG 62
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 83 FPVEKDLIEAIRASNGETLEKINSRPPCVIL 115
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 63 FPVEDDIMGVIGRAYDQGPVEKITKSGPPCVIL 94
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 6
04T504_TETNG
ID 04T504_TETNG PRELIMINARY; PRT; 89 AA.
AC 04T504;
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF9476, whole genome shotgun sequence
   (Chromosome undetermined SCAF9556, whole genome shotgun
   sequence).
GN ORNames=GSTENG00007059001, GSTENG00007180001;
OS Tetradon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorphi; Tetradontiformes;
OC Tetradontidae; Tetradontidae; Tetradon.
NCBI_TaxID=99883;

```

RA RA NUCLEOTIDE SEQUENCE.
RA Tallion O., Aury J.M., Brunet F., Petic J.L., Strange-Thomann N.,
RA Mauceli E., Bonneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicoud S., Jaffe D., Fisher S., Lutali G., Dosset C., Segurens B.,
RA Dasilva C., Salenobad M., Levy M., Bouder N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Koltika M., Vacherie B.,
RA Biemont C., Skalli Z., Catcollco L., Poulan J., De Berardinis V.,
RA Craud C., Dupret S., Broctier P., Coutanceau J.P., Gouy J.,
RA Parra G., Landrieu G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Keills M., Volff J.N., Guigro R., Zody M.C., Meikrov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Lander V., Schachter V., Querier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.,
RT "Genome duplication in the teleost fish Tetradodon nigrovittatus reveals
RL the early vertebrate proto-karyotype." ;
RN Nature 431:946-957(2004).
RP [2]
RR NUCLEOTIDE SEQUENCE.
RG Genoscope, Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.
RC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CD EMBL, CA801009456; CAP92028.1; -; Genomic DNA.
DR EMBL, CA801009556; CAF92112.1; -; Genomic DNA.
DR EMBL, CA801009556; CAF92112.1; -; Genomic DNA.
SQ SEQUENCE 89 AA; 9818 MW; 1F53DB9D7B886CA CRC64;

Query Match 36.5%; Score 218; DB 2; Length 89;
Best Local Similarity 51.2%; Pred. No. 1,3e-11;
Matches 42; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

OY 34 GFATVYELASAVKEQYGGIEBSRLGCTGAFFIEINGVLFSKLENGGPPYEKOLIBAI 93
Db 8 GYGGRYVELLRVVAKAFDLADVGEGVGLTSGFEIVINEGVFTSLTETGFGPEEDVDLVQVI 67

OY 94 RRASNGETLETINSPRECVIL 115
Db 68 QCAVDGKPEVKLTITSRPPCVM 89

RESULT 7
O9VRAO_DROME PRELIMINARY; PRT; 95 AA.

AC O9VRAO; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE CG15456-PA.
GN Name=CG15456; ORFNames=CG15456;
OS Drosophila melanogaster (Fruit fly).
OC Neoptera; Endopterygota; Hexapoda; Insecta; Pterygota;
OC Ephemeroptera; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
[1]
NP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Gelinkler S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galley R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Abghyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balcer R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Chew J.S., Dahlke S., Davidson L.B., Davies P.,
RA de Pablos B., Delcher A., Deng X., Davis A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuseken D.R., Pacleb J.M.,
RA Palenzola M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Ralner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Sider B.C., Sider-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spter E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of *Drosophila melanogaster*,"
RA Science 287:2185-2195(2000).
[2]
RA NUCLEOTIDE SEQUENCE.
RA MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RA "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RA melanogaster euchromatic genome sequence."
RA Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RA NUCLEOTIDE SEQUENCE.
RA MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.,
RA "The transposable elements of the *Drosophila melanogaster* euchromatin:
RA a genomic perspective."
RA Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
[4]
RA NUCLEOTIDE SEQUENCE.
RA MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tuzy J.L., Whitefield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RA systematic review."
RA Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
[5]
RA NUCLEOTIDE SEQUENCE.
RA Berkeley *Drosophila* Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirkas R., Smith E.,
RA Yu C., Rubin G.,
RA "Drosophila melanogaster release 4 sequence."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[6]
RA NUCLEOTIDE SEQUENCE.
RA FlyBase;
RA Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC - INTERACTION:
CC OBTAIN:CG15056; NDBEXP-1; Inact=EBI-182672, EBI-135162;
DR EMBL/ A003571; AAF50903.1; -, Genomic_DNA.
DR Inact; Q9VRA0; -.

DR Ensembl; CG15456; *Drosophila melanogaster*.
DR FlyBase; FBgn0040650; CG15456.
DR GO; GO:0005739; C-mitochondrion; IEA.
DR GO; GO:0005489; F-electron transporter activity; IEA.
DR GO; GO:0005506; F:ion ion binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR011893; CXXU_selWTH.
DR InterPro; IPR010041; Ferredoxin.
DR TrEMBL; TIGR02174; CXXU_selWTH; 1.
SQ SSQUENCE 95 AA; 10506 MW; A4564893BF263FDS CRC64;
Query Match 24.9%; Score 148.5; DB 2; Length 95;
Best Local Similarity 34.0%; Pred. No. 9.6e-07;
Matches 32; Conservative 17; Mismatches 44; Indels 1; Gaps 1;
QY 23 VRIVEYCEPCGEATYLELASVKEQVPEIESRLGTGAPEIEINGQLVPSKLENG 82
DB 2 VKVEVEYCGICNFSGGCHLREFLASSPDLISCRTGRGSPFVSIDGQVHASKLSCLA 61
QY 83 FPEYKDLIAIRASNGETLEKTNR-PPCVIL 115
DB 62 FPGHSAVLAVQVAKERGEPEYKYLEQPIKCVVM 95
RESULT 8
ID Q8H6T4_CHLRE PRELIMINARY; PRT; 88 AA.
AC Q8H6T4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Selenoprotein selW.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonadae.
OX NCBI_Taxid=3055;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Novoselov S.V., Rao M., Onoshko N.V., Zhi H., Kryukov G.V., Xiang Y.,
RA Weeks D.P., Hatfield D.L., Gladyshev V.N.,
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF494050; AAN32901.1; -, mRNA.
DR InterPro; IPR011893; CXXU_selWTH.
DR TrEMBL; TIGR02174; CXXU_selWTH; 1.
KW Selenium; Selenocysteine.
FT SE CVS 14
KW 14
SQ SSQUENCE 88 AA; 9690 MW; DF27CBA4780E1128 CRC64;
Query Match 23.0%; Score 137.5; DB 2; Length 88;
Best Local Similarity 37.2%; Pred. No. 1.1e-05;
Matches 32; Conservative 17; Mismatches 32; Indels 5; Gaps 2;
QY 23 VRIVEYCEPCGEATYLELASVKEQVPEIEI-----ESRLGTGAPEIEINGQLVPSK 78
DB 4 VQVHVLVCGCGGSGSRYSLENAIRMKFPNADIKFSEFATPPQATGFEVNVNELVHASK 63
QY 79 ENGPFYKDLIAIRASNGETLEK 104
DB 64 NGGAVHDNOEKVERI-FAKIGERLAK 88
RESULT 9
Q4UBM6_SULAC
ID Q4UBM6_SULAC PRELIMINARY; PRT; 80 AA.
AC Q4UBM6;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
DE OrderedLocName=Sac1_0387;
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC Sulfolobus.
 OX NCBI_TaxID=2285;
 RN [1]
 RC NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RP STRAIN-ATCC 33909 / NCIB 11770 / DSM 639.
 RX PubMed:15955215; DOI=10.1128/JB.187.14.4992-4999.2005;
 RA Chen L., Bruegger K., Skovgaard M., Redder P., She Q., Torarinsson E.,
 RA Greve B., Aweyer M., Zibat A., Klenk H.-P., Garrett R.A.;
 RT "The genome of *Sulfolobus acidocaldarius*, a model organism of the
 RT Crenarchaeota.";
 RL J. Bacteriol. 187:4992-4999(2005).
 DR EMBL; CP000077; AAY79803.1; -; Genomic_DNA.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 80 AA; 9219 MW; 607FF3E7589F304 CRC64;
 Query Match 22.9%; Score 137; DB 2; Length 80;
 Best Local Similarity 31.2%; Pred. No. 1.1e-05;
 Matches 25; Conservative 21; Mismatches 32; Indels 2; Gaps 1;
 QY 21 SGVRIVYCEPCGEATYLELASAVKEQYPGIEISRLGGTGAPEIRINGQLVFSKLEN 80
 Db 3 TNNKIV-YCRPCGLDRLANLARDLSYEGVNELEGGKNGIFDYVDQLIFSRKE 60
 QY 81 GGPEYKDLIAIRASNGE 100
 Db 61 KRFPDGEILKELSKKATAQ 80
 RESULT 10
 Q7QFL8 ANOGA PRELIMINARY; - PRT; 93 AA.
 ID Q7QFL8 ANOGA PRELIMINARY; - PRT; 93 AA.
 AC Q7QFL8;
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE ENSANGP0000007373 (Fragment).
 GN ORFNames=ENSANG00000005563;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygora;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;
 OC Anophelinae; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=PEST;
 RG The Anopheles gambiae Sequence Committee;
 RT "Anopheles gambiae re-annotation.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=PEST;
 RG The Anopheles gambiae Sequence Committee;
 CC Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008846; EA06627.2; -; Genomic_DNA.
 DR InterPro; IPR011893; CXU_selWTH.
 DR TIGRFAMs; TIGR02174; CXU_selWTH; 1.
 FT NON TER 1
 SQ SEQUENCE 93 AA; 10351 MW; DC3808C7B37D169 CRC64;
 Query Match 22.2%; Score 132.5; DB 2; Length 93;
 Best Local Similarity 33.7%; Pred. No. 3.5e-05;
 Matches 28; Conservative 20; Mismatches 34; Indels 1; Gaps 1;
 QY 33 CGFATYLELASAVKEQYPGIEISRLGGTGAPEIRINGQLVFSKLENGPEYKDLIA 92
 Db 10 CNSKPCGLELALALREQLPELWVCRGCRGSRGFEVQINDTLVHSLGSLAFPRYEEVQN 69
 QY 93 IRRASNGETLEKITSNR-PCVIT 114
 Db 70 VRNARDGLFVARVAEQPTDCLV 92

RESULT 11
 O8S227 ORYSA PRELIMINARY; PRT; 232 AA.
 ID Q8S227 ORYSA PRELIMINARY; PRT; 232 AA.
 AC Q8S227;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-FEB-2005 (TREMBlrel. 29, Last annotation update)
 DE Hypothetical protein P0446G04.43-2 (Hypothetical protein P0460C04.2-2).
 DE 2).
 GN Name=P0446G04.43-2; Synonyms=P0460C04.2-2;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
 RA Wu J., Nishimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
 RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
 RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
 RA Hijiishi S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
 RA Ikono M., Itoh S., Itoh T., Itoh Y., Iwabuchi A., Kamiya K.,
 RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
 RA Machida K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
 RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
 RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
 RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
 RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
 RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
 RA Yano M., Jiang J., Gojobori T.;
 RT "The genome sequence and structure of rice chromosome 1.";
 RL Nature 420:312-316(2002).
 DR EMBL; AP003252; BAB89601.1; -; Genomic_DNA.
 DR EMBL; AP003252; BAB89601.1; -; Genomic_DNA.
 DR Gramene; Q8S227; -;
 DR InterPro; IPR011893; CXU_selWTH.
 DR TIGRFAMs; TIGR02174; CXU_selWTH; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 232 AA; 25385 MW; 7011E960E7909E4F CRC64;
 Query Match 19.9%; Score 119; DB 2; Length 232;
 Best Local Similarity 20.7%; Pred. No. 0.0021;
 Matches 34; Conservative 19; Mismatches 39; Indels 72; Gaps 2;
 QY 14 PEEVEPGSGVRIYVCEPCGEATYLELASAVKEQYPGIEI-----ESRLGGTGA 55
 Db 67 PVDGPRASGTVLALKFCASCSYNGNAVTKMKLETSPGIVHVLNYPFPKRLSKAV 126
 QY 56 -----
 Db 127 PFLQVGMATLADGDIFFRFNGVPPWVYSLRANRGTMATLWLFNFAQSLQSSGAF 186
 QY 66 EIRINGQLVFSKLENGPEYKDLIAIR-----ASNGETLEKI 105
 Db 187 EVTCNGQLVFSKLSBOQFSEFLRELIGNRLPDSQPGKLEKY 230
 RESULT 12
 HS66 HETGL STANDARD; PRT; 244 AA.
 ID HS66 HETGL STANDARD; PRT; 244 AA.
 AC Q8BN19;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Putative esophageal gland cell secretory protein 6 precursor.
 GN Name=HS66;
 OS Heterodera glycines (soybean cyst nematode worm).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
 OX NCBI_TaxID=51029;

RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=2104659; PubMed=11310741;
 RA Wang X., Allen R., Ding X., Goellner M., Maier T., de Boer J.M.,
 RA Baum T.J., Hasey R.S., Davis B.L.,
 RT "Signal peptide-selection of cDNA cloned directly from the esophageal
 gland cells of the soybean cyst nematode *Heterodera glycines*.";
 RL Mol. Plant Microbe Interact. 14:536-544(2001).
 CC -1- SIMILARITY: Belongs to the SELT family.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL, AF273733, AAC21336.2, -, mRNA.
 DR InterPro: IPR011893; CXXU_selWTH.
 DR Trifurcates; TIGR02174; CXXU_selWTH; 1.
 KM Hypothetical protein; Redox-active center; Signal.
 FT SIGNAL 1 22 Potential.
 FT CHAIN 23 244 Putative esophageal gland cell secretory
 FT DISUFID 88 91 Redox-active (Potential).
 FT SEQUENCE 244 AA; 27201 MW; DD94A7A590AA9143 CRC64;
 SQ
 Query Match 19.2%; Score 114.5; DB 1; Length 244;
 Best Local Similarity 22.5%; Pred. No. 0.0063;
 Matches 39; Conservative 15; Mismatches 30; Indels 89; Gaps 5;
 QY 15 EEV---EPGSGVR-----IVEYCEPCGEATYELASAVKQYRGIE----- 54
 DB 60 EEWIRRP-SGTKSRFLPIMPVPKSPFCVSCGYRQAYEPQALKEKPGIDHENV 118
 QY 55 ----- 54
 DB 119 PGILRTVGAOVIGMKIALIVCVSGSPPTLGLTFPFQMLSNRLSALMLFLFS 178
 QY 55 --IESRLGCGAPRIENGOLVPSKENGSPRYKDLIE-----AIRPSNG 99
 DB 179 NALGMLQSTGAPEIYESIRKWSLKGVRPPELFOALDSHLAIRGAG 231
 RESULT 13
 SEPM1 HUMAN STANDARD; PRT; 86 AA.
 ID SEPM1_HUMAN O15532; O19096; Q86T19; Q96K05;
 AC P63302; O15532; O19096; Q86T19; Q96K05;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Selenoprotein W.
 GN Name=SEPM1; Synonyms=SELW;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
 OC Homo.
 NC NCBL_TaxID=9606;
 OK NCBL_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Placenta;
 RA Bellingham J., Gregory-Evans C.Y.,
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Sketleal muscle;
 RX MEDLINE=733815; PubMed=9256076; DOI=10.1016/S0378-1119(97)00113-3;
 RA Gu Q.-P., Bellarein M.A., Vendeland S.C., Lugade A., Ream W.,
 RA Whanger P.D.,
 RT "Conserved features of selenocysteine insertion sequence (SECIS)
 RT elements in selenoprotein W cDNAs from five species.";
 RL Gene 193:187-196(1997).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.

RA Bellingham J.,
 RT "Genomic structure of human selenoprotein W (SEPM1).";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Brain, PNS, Testis, and Uterus;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603699;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
 RA Diatchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Sapletone M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schenck A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May be involved in a redox-related process. May play a
 CC role in the myopathies of selenium deficiency (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL, AF015283, AAB69859.1, -, mRNA.
 DR EMBL, AF015283; AAC51665.1; -, mRNA.
 DR EMBL, AF247455; AL000898.1; -, Genomic DNA.
 DR EMBL, AF247454; AL000898.1; JOINED; Genomic DNA.
 DR EMBL, BC000581; AA000581.1; -, mRNA.
 DR EMBL, BC032546; AA032546.1; -, mRNA.
 DR EMBL, BC039597; AA039597.1; -, mRNA.
 DR EMBL, BC047893; AA047893.2; -, mRNA.
 DR EMBL, ENSG00000178980; Homo sapiens.
 DR HGN: HGN:10752; SEPM1.
 DR MIM: 603235; -
 DR GO: GO:0005737; C:cytoplasm; NAS.
 DR GO: GO:0005886; C:plasma membrane; ISS.
 DR GO: GO:0016491; P:oxidoreductase activity; ISS.
 DR InterPro: IPR011893; CXXU_selWTH.
 DR Trifurcates; TIGR02174; CXXU_selWTH; 1.
 KM Redox-active center; Selenium; Selenocysteine.
 FT INIT MET 0 0 By similarity.
 FT SE CYS 12 12
 FT DISUFID 9 12 Redox-active (By similarity).
 FT SEQUENCE 86 AA; 9270 MW; 21297DBD1815F7B8 CRC64;
 SQ
 Query Match 18.8%; Score 112.5; DB 1; Length 86;
 Best Local Similarity 33.3%; Pred. No. 0.003;
 Matches 27; Conservative 21; Mismatches 24; Indels 9; Gaps 4;
 QY 23 VRIIVEYCEPCGEATYELASAVKQYRG-IT--BSRLGCGAPRIENGOLVPSKLE 79
 DB 4 VRIIV--YCGACGCKSYLQLKKLEDFPRLDICSGTPOAGTFEWWAGKLHKKK 61
 QY 80 NGGFPYEKD-----LIRAIRA 96
 DB 62 GGGIVDTSEKFLKLVAAIKKA 82
 RESULT 14

```

SEPM1 MACMU
ID SEPM1 MACMU STANDARD; PRT; 86 AA.
AC P63303; 015532; 019096; 086T19; Q36KMS;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Selenoprotein W.
GN Name=SEPM1; Synonyms=SELM;
OS Macaca mulatta (rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
OC Cercopithecoidea; Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=97398151; PubMed=9256076; DOI=10.1016/S0378-1119(97)00113-3;
RA Gu Q.-Y., Bellstein M.A., Vendeland S.C., Lugade A., Ream W.,
RA Whanger P.D.;
RT "Conserved features of selenocysteine insertion sequence (SECIS)
RL elements in selenoprotein W cDNAs from five species.";
CC -1- FUNCTION: May be involved in a redox-related process. May play a
CC role in the myopathies of selenium deficiency (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U67450; AAC51666.1; -, mRNA.
DR GO; GO:0005739; C:mitochondrion; NAS.
DR GO; GO:0005886; C:plasma membrane; ISS.
DR GO; GO:0003954; F:NADH dehydrogenase activity; NAS.
DR GO; GO:0016491; F:oxidoreductase activity; ISS.
DR InterPro; IPR011893; CXXU_selWTH; 1.
DR TIGRFAMs; TIGR02174; CXXU_selWTH; 1.
KW Redox-active center; Selenium; Selenocysteine.
FT INIT MET 0
FT SE CYS 12
FT DISULFID 9 12
FT SEQUENCE 86 AA; 9270 MW; 21297DBD1815F7B8 CRC64;
SQ
Query Match 18.8%; Score 112.5; DB 1; Length 86;
Best Local Similarity 33.3%; Pred. No. 0.003;
Matches 27; Conservative 21; Mismatches 24; Indels 9; Gaps 4;
OY 23 VRIVVEGCEPGCEATYLELASAVKEQPG-IEI--ESRLGCTGAPEIEINGOLVFSEKL 79
DB 4 VRVV--YCGAGGYKSKYQLQKKLEDFPGRLDICGEGTPQATGFEVAVAGKLVHSSKK 61
OY 80 NGGFPYEKD---LIEAIRRA 96
DB 62 GGDGYVDTESKFLKLVAAIKAA 82

```

```

RESULT 15
SEPM1 PIG STANDARD; PRT; 86 AA.
AC Q9SKL4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Selenoprotein W.
GN Name=SEPM1; Synonyms=SEPM;
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OX NCBI_TaxID=9823;
RN [1]

```

```

RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Smooth muscle;
RA Costra N.D., Vorchek W.R., Whanger P.D.;
RT "Amino acid sequence of porcine selenoprotein W derived from analysis
RL of a cDNA library prepared from porcine smooth muscle.";
RT Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: May be involved in a redox-related process. May play a
CC role in the myopathies of selenium deficiency (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; AF301118; AAK56322.1; -, mRNA.
DR GO; GO:0005737; C:cytoplasm; ISS.
DR GO; GO:0005886; C:plasma membrane; ISS.
DR GO; GO:0016491; F:oxidoreductase activity; ISS.
DR InterPro; IPR011893; CXXU_selWTH; 1.
DR TIGRFAMs; TIGR02174; CXXU_selWTH; 1.
KW Redox-active center; Selenium; Selenocysteine.
FT INIT MET 0
FT SE CYS 12
FT DISULFID 9 12
FT SEQUENCE 86 AA; 9167 MW; 20C80134ABFRAAD22 CRC64;
SQ
Query Match 18.8%; Score 112.5; DB 1; Length 86;
Best Local Similarity 34.1%; Pred. No. 0.003;
Matches 28; Conservative 19; Mismatches 28; Indels 7; Gaps 3;
OY 22 GVRIVVEGCEPGCEATYLELASAVKEQPG-IEI--ESRLGCTGAPEIEINGOLVFSEKL 78
DB 1 GVAARVVYCGAGGYKSKYQLQKKLEDFPGRLDICGEGTPQATGFEVAVAGKLVHSSKK 60
OY 79 NGGFPYEKD---LIEAIRRA 96
DB 61 GGDGYVDTESKFLKLVAAIKAA 82

```

Search completed: March 27, 2006, 12:07:31
Job time : 139.659 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2006, 12:07:54 ; Search time 29.9551 Seconds
(without alignments)
317.399 Million cell updates/sec

Title: US-09-824-787B-2

Perfect score: 597

Sequence: 1 MSSEPCQTSVAPPEPEVEPG.....ASNGETLEKTNRPCCVIL 115

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/H.COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/RE.COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403	67.5	90	2 US-09-513-999C-7234	Sequence 7234, App1
2	96.5	16.2	88	1 US-08-807-043-3	Sequence 3, App1
3	96.5	16.2	88	1 US-09-127-289-3	Sequence 3, App1
4	96.5	16.2	88	2 US-09-393-448-3	Sequence 3, App1
5	96.5	16.2	93	1 US-08-807-043-1	Sequence 1, App1
6	96.5	16.2	93	1 US-09-127-289-1	Sequence 1, App1
7	96.5	16.2	93	2 US-09-393-448-1	Sequence 1, App1
8	87	14.6	216	2 US-09-270-767-46297	Sequence 46297, A
9	83	13.9	99	2 US-09-252-991A-17867	Sequence 17867, A
10	83	13.9	119	2 US-09-489-039A-10006	Sequence 10006, A
11	72	12.1	352	2 US-09-248-796A-17066	Sequence 17066, A
12	71.5	12.0	400	2 US-09-252-991A-22821	Sequence 22821, A
13	70	11.7	695	2 US-09-489-039A-14338	Sequence 14338, A
14	69	11.6	874	2 US-10-163-214-13	Sequence 13, App1
15	69	11.6	915	2 US-10-163-214-6	Sequence 6, App1
16	68.5	11.5	409	2 US-09-252-991A-29381	Sequence 29381, A
17	68	11.4	417	2 US-09-433-241A-10	Sequence 10, App1
18	68	11.4	735	2 US-09-115-704-2	Sequence 2, App1
19	68	11.4	735	2 US-09-780-115-2	Sequence 2, App1
20	68	11.4	909	2 US-10-163-214-2	Sequence 2, App1
21	67.5	11.3	1372	2 US-09-902-540-14099	Sequence 14099, A
22	67	11.2	299	2 US-09-720-318A-4	Sequence 4, App1
23	67	11.2	696	1 US-08-765-081-5	Sequence 5, App1
24	67	11.2	696	2 US-09-098-082-5	Sequence 5, App1
25	67	11.2	696	2 US-09-098-082-5	Sequence 5, App1
26	67	11.2	703	4 PCT-US95-06994-8	Sequence 8, App1
27	67	11.2	718	4 PCT-US95-06994-6	Sequence 6, App1

28	66.5	11.1	282	2 US-09-198-452A-739	Sequence 739, App
29	66.5	11.1	287	2 US-09-438-185A-699	Sequence 699, App
30	66.5	11.1	772	2 US-09-907-794A-339	Sequence 339, App
31	66.5	11.1	772	2 US-09-905-125A-339	Sequence 339, App
32	66.5	11.1	772	2 US-09-902-775A-339	Sequence 339, App
33	66.5	11.1	772	2 US-09-906-700-339	Sequence 339, App
34	66.5	11.1	772	2 US-09-903-603A-339	Sequence 339, App
35	66.5	11.1	772	2 US-09-904-920A-339	Sequence 339, App
36	66.5	11.1	772	2 US-09-909-064-339	Sequence 339, App
37	66.5	11.1	772	2 US-09-905-381A-339	Sequence 339, App
38	66.5	11.1	772	2 US-09-906-618-339	Sequence 339, App
39	66.5	11.1	772	2 US-09-906-646-339	Sequence 339, App
40	66.5	11.1	772	2 US-09-904-462-339	Sequence 339, App
41	66.5	11.1	772	2 US-09-902-736A-339	Sequence 339, App
42	66.5	11.1	772	2 US-09-906-722A-339	Sequence 339, App
43	66.5	11.1	2675	2 US-09-477-962-101	Sequence 101, App
44	66	11.1	344	2 US-09-489-039A-7249	Sequence 7249, App
45	66	11.1	393	2 US-09-540-236-3416	Sequence 3416, App

ALIGNMENTS

```

RESULT 1
US-09-513-999C-7234
; Sequence 7234, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclet, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7234
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 18
; OTHER INFORMATION: Xaa=Asp or Glu
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 46
; OTHER INFORMATION: Xaa=Leu or Val
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 78
; OTHER INFORMATION: Xaa=Ala or Gly
US-09-513-999C-7234

Query Match      67.5%; Score 403; DB 2; Length 90;
Best Local Similarity 87.6%; Pred. No. 4.1e-40;
Matches 78; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSSEPCQTSVAPPEPEVEPGSGRIVYECPCGFEATYLEASAKVEQYPGIEISRLG 60
    |||||
DB 1 MSSEPCQTSVAPPEPEVEPGSGRIVYECPCGFEATYLEASAKVEQYPGIEISRLG 60
    |||||

QY 61 GTGAFIEIINGQLVFSKLENGSPPEYKDL 89
    |||||
DB 61 GTGAFIEIINGQLVFSKLENGSPPEYKDL 89
    |||||

RESULT 2
US-08-807-043-3

```

```

Sequence 3, Application US/08807043
Patent No. 5856131
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN SELENOPROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,043
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0202 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 88 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 993035
US-08-807-043-3
Query Match      16.2%; Score 96.5; DB 1; Length 88;
Best Local Similarity   34.1%; Pred. No. 0.00087;
Matches    28; Conservative   17; Mismatches    26; Indels   11; Gaps     5
QY          23 VRIVERCEPGCFEATYLELSAVKEQYPG-IEI--ESRLGGTGAFFELINGQLVFSKLE 79
|||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          5 VRVVV-YCGAGGYRKYQLQTKRKLEHPPGCIDIGEGTPGYTGPFEEVYAGKLVAHSKR 62
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY          80 NGGF-----PYEKDLEAIRRA 96
|-:-:-:-|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          63 GDGVDTESKFRK-LVTAIKAA 83
|-:-:-:-|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 3
US-09-127-289-3
Sequence 3, Application US/09127289
Patent No. 5998371
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN SELENOPROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA

```

```

? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/127,289
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/807,043
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0202 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? TELEX:
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 88 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: GenBank
? CLONE: 993035
? US-09-127-289-3

Query Match 16.2%; Score 96.5; DB 1; Length 88;
Best Local Similarity 34.1%; Pred. No. 0.00087;
Matches 28; Conservative 17; Mismatches 26; Indels 11; Gaps 5.

QY 23 VRVVECEPGCFATYELASAVKEGYPG-IET--ESRLGGTGAPEIENGOLVFSTLE 79
DB 5 VRRV--YCGAGXGYPKYLQLKKEKHHEFPGLDLCGETSPQVTGTFEYTVAGKLVSKKR 62
      ||| |::||::||::||::||::||::||::||::||::||::||::||::||
      .|. :.|:|. :.|:|. :.|:|. :.|:|. :.|:|. :.|:|. :.|:|. :.|
QY 80 NGGF----PYEKDLIEAIRRA 96
      .|. :.|:|. :.|:|. :.|:|. :.|:|. :.|:|. :.|:|. :.|:|. :.|
DB 63 GDGVVDTSKFRK-LVTAIKAA 83

RESULT 4
US-09-393-448-3
Sequence 3, Application US/09393448
Patent No. 6545129
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN SELENOPROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/393,448
FILING DATE: 10-Sep-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,043
FILING DATE: <Unknown>

```

```

; INFORMATION FOR SEQ ID NO:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 93 amino acids
;     TYPE: amino acid

```

23 VRIVVEYCEPCGEATYLELASAVKQYPG-IEI--ESRLGTGAFEIEINGOLVFSKLE 79

Db 11 VRV--YCGAGYKSKYQLKKKLEDPGRDLICGEGTSQAXGFPEVWVAGKLHSHKXK 68
QY 80 NGGF 83
Db 69 GDGY 72

RESULT 7

US-09-393-448-1
Sequence 1, Application US/09393448
Patent No. 6545129

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN SELENOPROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/393,448
FILING DATE: 10-Sep-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/807,043
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0202 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BLADNOT03
CLONE: 1599862

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-393-448-1

Query Match 16.2%; Score 96.5; DB 2; Length 93;
Best Local Similarity 32.8%; Pred. No. 0.00094;
Matches 21; Conservative 18; Mismatches 20; Indels 5; Gaps 3;

QY 23 VRIVEYCEPCGFATYLELASAVKEQYPG-IEI--ESRLGGTGAPEIEINGQLVFSKLE 79
Db 11 VRV--YCGAGYKSKYQLKKKLEDPGRDLICGEGTSQAXGFPEVWVAGKLHSHKXK 68

QY 80 NGGF 83
Db 69 GDGY 72

RESULT 8

US-09-270-767-46297
Sequence 46297, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 46297
LENGTH: 216
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-46297

Query Match 14.6%; Score 87; DB 2; Length 216;
Best Local Similarity 18.8%; Pred. No. 0.042;
Matches 25; Conservative 15; Mismatches 25; Indels 68; Gaps 1;

QY 29 YCEPCGFATYLELASAVKEQYPGIEI-----ESRLGGTGAPEIEINGQLVFSKLEN 80
Db 66 YCSCGYKAFEDYVGLGEXYPOIVGNGVDPGILNYLSKMIFALKIIVSVSAV 125

QY 56 -----
Db 126 SPFTFLGLNTPSWSHQANKIYACMMIFLGNMLEAQLISSGAPEITLNDVPVMSKLOT 185

QY 81 GGFPEYKDLIEAI 93
Db 186 GRFSPPEVLFOII 198

RESULT 9

US-09-252-991A-17867
Sequence 17867, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 17867

LENGTH: 99

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17867

Query Match 13.9%; Score 83; DB 2; Length 99;
Best Local Similarity 31.2%; Pred. No. 0.041;
Matches 25; Conservative 13; Mismatches 36; Indels 6; Gaps 3;

QY 19 PGSGRIYVCEPCGF--EATYL--ELASAVKEQYPGIEISRLGGTGAPEIEINGQLV 74
Db 5 PTAKEPIVITCTCCOMLTRAMLAQELSTFADDLKVCLEPGTG--GVRRITCDGVQV 62

QY 75 FSKLNGGFPYKDLIEAIR 94
Db 63 WERKADGQFPEAKALKQVR 82

RESULT 10

US-09-489-039A-10006
Sequence 10006, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 10006
LENGTH: 119
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10006

Query Match 13.9%; Score 83; DB 2; Length 119;
Best Local Similarity 25.3%; Pred. No. 0.054;
Matches 21; Conservative 17; Mismatches 39; Indels 6; Gaps 3;

QY 16 EVERSGRIVVEYCEPCGF--EATYL--ELASAVEQYPGIEISRLGGTGAFFIEING 71
DB 20 ETWNSKALITITYSQCWMLRASMMAOELHTEFTDIASVTLVPTG--GIFDIVDG 77
QY 72 QLVFSKLENGGFPYKDLIEAIR 94
DB 78 QQWERKODGGPDAALRRVR 100

RESULT 11
US-09-248-796A-17066
Sequence 17066, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 17066
LENGTH: 352
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-17066

Query Match 12.1%; Score 72; DB 2; Length 352;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 21; Conservative 14; Mismatches 32; Indels 18; Gaps 2;

QY 33 CGFATYIELASAVEQYPGIEISRLGGTGAFFIEINGQLVFSKLE--NG----- 82
DB 158 CAGTVVYKALTAELIAGQWVAISGAGLSLAVQYAKMGVRYTLAIDGEDKGEFVKS 217
QY 83 -----FPYKXLEAIRRASNG 99
DB 218 LGAEFTIDFTYKQVAVKKAATNG 242

RESULT 12
US-09-252-991A-22821
Sequence 22821, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,768
PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22821
LENGTH: 400
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22821

Query Match 12.0%; Score 71.5; DB 2; Length 400;
Best Local Similarity 29.5%; Pred. No. 7.1;
Matches 26; Conservative 10; Mismatches 33; Indels 19; Gaps 4;

QY 5 PGQTSVAPPEPEVP-----GSGVRIYVEYCEPCGFATYLE--LASAVEQYPG 52
DB 297 PATSPAPAPAPSEPPAAPVAVAGGQV--VYQVADCTQVTDANGVYVSAKRGDS 355
QY 53 IE-----IESRLGGTGAFFIEINGQLV 74
DB 356 LELAGKAPLELRIGFARGAQVSYNGQPV 383

RESULT 13
US-09-489-039A-14338
Sequence 14338, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 14338
LENGTH: 695
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14338

Query Match 11.7%; Score 70; DB 2; Length 695;
Best Local Similarity 22.2%; Pred. No. 24;
Matches 18; Conservative 20; Mismatches 39; Indels 4; Gaps 1;

QY 29 YCEPCGFATYIELASAVEQYPGIEISRLGGTGAFFIEINGQLVFSKLENGGFPYKXD 88
DB 531 YSGVGMQT-----LSNAGKADATGVELEAKWRFAFGWSMDINGVIRSEFTNDSLEYHGN 586
QY 89 LIEAIRBASNGETLEKITSR 109
DB 587 RVFPVRYGAGSYNGVYIDTR 607

RESULT 14
US-10-163-214-13
Sequence 13, Application US/10163214
Patent No. 6849781
GENERAL INFORMATION:
APPLICANT: Allen, Stephen W.
APPLICANT: Broglie, Karen E.
APPLICANT: Butler, Karlene H.
APPLICANT: Thorpe, Catherine J.
TITLE OF INVENTION: Starch Synthase Isoform V
FILE REFERENCE: BB1520 US NA
CURRENT APPLICATION NUMBER: US/10/163,214
CURRENT FILING DATE: 2002-06-05
PRIOR APPLICATION NUMBER: 60/297,099
PRIOR FILING DATE: 2001-06-08
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Microsoft Office 97
SEQ ID NO 13

LENGTH: 874
TYPE: PRT
ORGANISM: Vigna unguiculata
US-10-163-214-13

Query Match 11.6%; Score 69; DB 2; Length 874;
Best Local Similarity 26.8%; Pred. No. 43;
Matches 22; Conservative 14; Mismatches 34; Indels 12; Gaps 3;

QY 21 SGVRIVVEYCEPCGFATYLELASAVKEQYPCGIEISRLGCTGAFELIINGQLVFSKLEN 80
DB 754 SDMFILPSIFEPG-----LTQMISMRYGALPIARKTGGLNDSVFDVDDTIFSQFRN 806

QY 81 GGFPY---EKDLIEAIRASN 98
DB 807 -GFTFLNADKGINDALVRAIN 827

RESULT 15

US-10-163-214-6
Sequence 6, Application US/10163214
Patent No. 6849781
GENERAL INFORMATION:
APPLICANT: Allen, Stephen M.
APPLICANT: Broglie, Karen E.
APPLICANT: Butler, Karlene H.
APPLICANT: Thorpe, Catherine J.
TITLE OF INVENTION: Starch Synthase Isoform V
FILE REFERENCE: BB1520 US NA
CURRENT APPLICATION NUMBER: US/10/163,214
PRIOR FILING DATE: 2002-06-05
PRIOR APPLICATION NUMBER: 60/297,099
PRIOR FILING DATE: 2001-06-08
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Microsoft Office 97
SEQ ID NO 6
LENGTH: 915
TYPE: PRT
ORGANISM: Oryza sativa
US-10-163-214-6

Query Match 11.6%; Score 69; DB 2; Length 915;
Best Local Similarity 28.4%; Pred. No. 46;
Matches 23; Conservative 11; Mismatches 37; Indels 10; Gaps 2;

QY 21 SGVRIVVEYCEPCGFATYLELASAVKEQYPCGIEISRLGCTGAFELIINGQLVFSKLEN 80
DB 797 SDMFILPSIFEPG-----LTQMISMRYGALPIARKTGGLNDSVFDVDDTIFSKELRN 849

QY 81 G--GFPYKDLIEAIRASN 98
DB 850 GFTFVHPDEKALSGAMERAFN 870

Search completed: March 27, 2006, 12:10:05
Job time : 30.9551 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2006, 12:32:34 / Search time 12.7395 Seconds
(without alignments)
266.247 Million cell updates/sec

Title: US-09-824-787b-2

Perfect score: 597 MSGPQGTSAVPPEEVEPG.....ASNGETLEKITSRPPCVIL 115

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 174695 seqs, 29494374 residues

Total number of hits satisfying chosen parameters: 174695

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: Published Applications_AA_New:
1: /SIDS5/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /SIDS5/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3: /SIDS5/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
4: /SIDS5/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
5: /SIDS5/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
6: /SIDS5/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
7: /SIDS5/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
8: /SIDS5/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	115	US-11-152-288-11	Sequence 11, Appl
2	76.5	12.8	170	US-11-096-568A-10110	Sequence 10110, A
3	76.5	12.8	186	US-11-096-568A-10109	Sequence 10109, A
4	76.5	12.8	198	US-11-096-568A-10108	Sequence 10108, A
5	76	12.7	283	US-11-096-568A-31233	Sequence 31233, A
6	76	12.7	390	US-11-096-568A-31232	Sequence 31232, A
7	76	12.7	395	US-11-096-568A-31231	Sequence 31231, A
8	69	11.6	253	US-11-096-568A-25229	Sequence 25229, A
9	69	11.6	253	US-11-096-568A-25228	Sequence 25228, A
10	69	11.6	415	US-11-096-568A-25227	Sequence 25227, A
11	68.5	11.5	186	US-11-096-568A-22856	Sequence 22856, A
12	68.5	11.5	209	US-11-096-568A-22855	Sequence 22855, A
13	68.5	11.5	234	US-11-096-568A-22854	Sequence 22854, A
14	68.5	11.5	329	US-10-510-386-36	Sequence 36, Appl
15	68.5	11.5	330	US-10-510-386-36	Sequence 36, Appl
16	67	11.2	228	US-11-082-389-24	Sequence 24, Appl
17	66.5	11.1	697	US-11-096-568A-28319	Sequence 28319, A
18	66.5	11.1	769	US-10-467-657-3280	Sequence 3280, Ap
19	66.5	11.1	876	US-11-096-568A-28318	Sequence 28318, A
20	66.5	11.1	884	US-11-096-568A-28317	Sequence 28317, A
21	66	11.1	254	US-08-978-360A-510	Sequence 510, App
22	66	11.1	259	US-11-096-568A-21634	Sequence 21634, A
23	66	11.1	596	US-11-152-903-2	Sequence 2, Appl1
24	66	11.1	596	US-11-152-903-4	Sequence 4, Appl1
25	66	11.1	596	US-11-152-903-6	Sequence 6, Appl1

26	66	11.1	596	US-11-152-903-8	Sequence 8, Appl1
27	66	11.1	596	US-11-152-903-10	Sequence 10, Appl1
28	66	11.1	596	US-11-152-903-12	Sequence 12, Appl1
29	66	11.1	1566	US-10-453-372-1190	Sequence 1190, Ap
30	65.5	11.0	290	US-11-096-568A-68847	Sequence 68847, Ap
31	65.5	11.0	324	US-11-087-099-9909	Sequence 9909, Ap
32	65.5	11.0	397	US-11-096-568A-6846	Sequence 6846, Ap
33	65.5	11.0	419	US-11-096-568A-6845	Sequence 6845, Ap
34	65.5	11.0	1943	US-11-122-396-5	Sequence 5, Appl1
35	65.5	11.0	2392	US-10-330-773-907	Sequence 907, App
36	65	10.9	2343	US-10-330-773-904	Sequence 904, App
37	64	10.7	222	US-11-076-164-10	Sequence 10, Appl1
38	64	10.7	344	US-11-087-099-4046	Sequence 4046, Ap
39	63.5	10.6	303	US-10-878-556A-74	Sequence 74, Appl1
40	63.5	10.6	528	US-11-087-099-7839	Sequence 7839, Ap
41	63	10.6	134	US-10-993-543-24	Sequence 24, Appl1
42	63	10.6	526	US-10-467-657-6754	Sequence 6754, Ap
43	63	10.6	756	US-10-954-468-15	Sequence 15, Appl1
44	63	10.6	756	US-10-954-468-27	Sequence 27, Appl1
45	63	10.6	758	US-10-954-468-28	Sequence 28, Appl1

ALIGNMENTS

```

RESULT 1
US-11-152-288-11
Sequence 11, Application US/11155288
Publication No. US2006008468A1
GENERAL INFORMATION:
APPLICANT: Chian, Chih-Sheng
APPLICANT: Simard, John J.L.
TITLE OF INVENTION: COMBINATIONS OF TUMOR-ASSOCIATED
TITLE OF INVENTION: ANTIGENS IN DIAGNOSTICS FOR VARIOUS TYPES OF CANCERS
FILE REFERENCE: NANAK.050A
CURRENT APPLICATION NUMBER: US/11/155,288
PRIOR FILING DATE: 2005-06-17
PRIOR APPLICATION NUMBER: 60/580,969
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11
LENGTH: 115
ORGANISM: Homo sapiens
US-11-152-288-11

Query Match      100.0%; Score 597; DB 7; Length 115;
Best Local Similarity 100.0%; Pred. No. 4.8e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 MSGEPGQTSVAPPEEVEPGSSGRIIVYECPCGFETYLEASAVYEQYPGIEISRLG 60
Db      1 MSGEPGQTSVAPPEEVEPGSSGRIIVYECPCGFETYLEASAVYEQYPGIEISRLG 60
        61 GTGAFIEIINGQVFSKLENGSPPEYKDLIEATIRASNGETLEKITSRPPCVIL 115
        61 GTGAFIEIINGQVFSKLENGSPPEYKDLIEATIRASNGETLEKITSRPPCVIL 115

RESULT 2
US-11-096-568A-10110
Sequence 10110, Application US/11096568A
Publication No. US20060084240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
PRIOR FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 10110

```

```
LENGTH: 170
TYPE: PRT
ORGANISM: Triticum aestivum
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(170)
OTHER INFORMATION: Ceres Seq. ID no. 13583721
US-11-096-568A-10110
```

```
Query Match
Best Local Similarity 12.8%; Score 76.5; DB 7; Length 170;
Matches 22; Conservative 11; Mismatches 34; Indels 3; Gaps 2;
```

```
QY 21 SGVRIVCEPC-GFEATYLELASAVKEQYPGIE--IESRLGTGAFIEINQGVFSK 77
DB 83 SKRIVVACTQCQCFKRRLKVKEDLSAVPGVSVTINPEKRRRCLEIREGGDVFTS 142
QY 78 LENGFPYEX 87
DB 143 LQNMPPRFCK 152
```

RESULT 3

```
US-11-096-568A-10109
Sequence 10109, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 10109
LENGTH: 186
TYPE: PRT
ORGANISM: Triticum aestivum
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(186)
OTHER INFORMATION: Ceres Seq. ID no. 13583720
US-11-096-568A-10109
```

```
Query Match
Best Local Similarity 12.8%; Score 76.5; DB 7; Length 186;
Matches 22; Conservative 11; Mismatches 34; Indels 3; Gaps 2;
```

```
QY 21 SGVRIVCEPC-GFEATYLELASAVKEQYPGIE--IESRLGTGAFIEINQGVFSK 77
DB 99 SKRIVVACTQCQCFKRRLKVKEDLSAVPGVSVTINPEKRRRCLEIREGGDVFTS 158
QY 78 LENGFPYEX 87
DB 159 LQNMPPRFCK 168
```

RESULT 4

```
US-11-096-568A-10108
Sequence 10108, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 10108
LENGTH: 198
TYPE: PRT
ORGANISM: Triticum aestivum
```

```
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(198)
OTHER INFORMATION: Ceres Seq. ID no. 13583719
US-11-096-568A-10108
```

```
Query Match
Best Local Similarity 12.8%; Score 76.5; DB 7; Length 198;
Matches 22; Conservative 11; Mismatches 34; Indels 3; Gaps 2;
```

```
QY 21 SGVRIVCEPC-GFEATYLELASAVKEQYPGIE--IESRLGTGAFIEINQGVFSK 77
DB 111 SKRIVVACTQCQCFKRRLKVKEDLSAVPGVSVTINPEKRRRCLEIREGGDVFTS 170
QY 78 LENGFPYEX 87
DB 171 LQNMPPRFCK 180
```

RESULT 5

```
US-11-096-568A-31233
Sequence 31233, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 31233
LENGTH: 283
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(283)
OTHER INFORMATION: Ceres Seq. ID no. 13579073
US-11-096-568A-31233
```

```
Query Match
Best Local Similarity 12.7%; Score 76; DB 7; Length 283;
Matches 31; Conservative 16; Mismatches 36; Indels 50; Gaps 5;
```

```
QY 13 PPEVIEP-----GSGRIVVCEPCGFEATYLELASAVKEQYPGIE--IESRL-- 59
DB 96 PPNKRPYDLVGAALFGDGAAYIIGADPRECEAFPMELHYAVOQFLPGTONVIDGRUTE 155
QY 60 -----GGTGAFEIEN-----GOLVFSKLENGGF 83
DB 156 EGINFKLRDLPQKLEENIEEFCKLMGKAGDESMEFMDWVAHPPGPAIINRLT-KL 214
QY 84 PYEKDLIEAIRRA 96
DB 215 KLEKELKLESSRRA 227
```

RESULT 6

```
US-11-096-568A-31232
Sequence 31232, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 31232
LENGTH: 390
TYPE: PRT
```



```

; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 25227
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(415)
; OTHER INFORMATION: Ceres Seq. ID no. 12566621
US-11-096-568A-25227

Query Match      11.6%; Score 69; DB 7; Length 415;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 29; Conservative 17; Mismatches 45; Indels 30; Gaps 5;

OY 12 PPPEVEPGSGVRIY-----VEYCE-----PCGFATYLELASAVKEQYPGIEISR 58
DB 185 PPPDLEPLNTLALAYASMGVDVYVLTSDRDLPDGGSHFAQYVAKELKPGILVE-- 242
OY 59 LGGTGAFEIEINGOL-VPSKLENGFPPEKDLIEAIR-----RASNGETLEKITS 108
DB 243 -----CLTSDPRGDELAISLSLNSGLDVAHNIETRSIQRVVRDPRAGYDQSLAVLGA 297
OY 109 R 109
DB 298 R 298

RESULT 11
US-11-096-568A-22856
; Sequence 22856, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22856
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(186)
; OTHER INFORMATION: Ceres Seq. ID no. 12410397
US-11-096-568A-22856

Query Match      11.5%; Score 68.5; DB 7; Length 186;
Best Local Similarity 28.6%; Pred. No. 3.9;
Matches 28; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

OY 3 GEPGQTSVAPPPEVE-----PGSGVRIYVYCEPCGFATYLELASAVKEQYPGIEISR 58
DB 77 GKKEAEAVAPAAEEDVDVAAGKRVIVACTOC---RQFKIRAKYKE-----DLESY 128
OY 59 LGGT-----GAFEIEINGOLVPSKLENGFPY 85
DB 129 VSGVSVIINPQKPRRGCLIEIREGGGVFISLNMRRPF 166

RESULT 12
US-11-096-568A-22855
; Sequence 22855, Application US/11096568A
; Publication No. US20060048240A1
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22855
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(209)
; OTHER INFORMATION: Ceres Seq. ID no. 12410396
US-11-096-568A-22855

Query Match      11.5%; Score 68.5; DB 7; Length 209;
Best Local Similarity 28.6%; Pred. No. 4.6;
Matches 28; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

OY 3 GEPGQTSVAPPPEVE-----PGSGVRIYVYCEPCGFATYLELASAVKEQYPGIEISR 58
DB 100 GKKEAEAVAPAAEEDVDVAAGKRVIVACTOC---RQFKIRAKYKE-----DLESY 151
OY 59 LGGT-----GAFEIEINGOLVPSKLENGFPY 85
DB 152 VSGVSVIINPQKPRRGCLIEIREGGGVFISLNMRRPF 189

RESULT 13
US-11-096-568A-22854
; Sequence 22854, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22854
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(234)
; OTHER INFORMATION: Ceres Seq. ID no. 12410395
US-11-096-568A-22854

Query Match      11.5%; Score 68.5; DB 7; Length 234;
Best Local Similarity 28.6%; Pred. No. 5.3;
Matches 28; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

OY 3 GEPGQTSVAPPPEVE-----PGSGVRIYVYCEPCGFATYLELASAVKEQYPGIEISR 58
DB 125 GKKEAEAVAPAAEEDVDVAAGKRVIVACTOC---RQFKIRAKYKE-----DLESY 176
OY 59 LGGT-----GAFEIEINGOLVPSKLENGFPY 85
DB 177 VSGVSVIINPQKPRRGCLIEIREGGGVFISLNMRRPF 214

RESULT 14
US-10-510-386-36
; Sequence 36, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
```

```

; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22855
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(209)
; OTHER INFORMATION: Ceres Seq. ID no. 12410396
US-11-096-568A-22855

Query Match      11.5%; Score 68.5; DB 7; Length 209;
Best Local Similarity 28.6%; Pred. No. 4.6;
Matches 28; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

OY 3 GEPGQTSVAPPPEVE-----PGSGVRIYVYCEPCGFATYLELASAVKEQYPGIEISR 58
DB 100 GKKEAEAVAPAAEEDVDVAAGKRVIVACTOC---RQFKIRAKYKE-----DLESY 151
OY 59 LGGT-----GAFEIEINGOLVPSKLENGFPY 85
DB 152 VSGVSVIINPQKPRRGCLIEIREGGGVFISLNMRRPF 189

RESULT 13
US-11-096-568A-22854
; Sequence 22854, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22854
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(234)
; OTHER INFORMATION: Ceres Seq. ID no. 12410395
US-11-096-568A-22854

Query Match      11.5%; Score 68.5; DB 7; Length 234;
Best Local Similarity 28.6%; Pred. No. 5.3;
Matches 28; Conservative 10; Mismatches 37; Indels 23; Gaps 4;

OY 3 GEPGQTSVAPPPEVE-----PGSGVRIYVYCEPCGFATYLELASAVKEQYPGIEISR 58
DB 125 GKKEAEAVAPAAEEDVDVAAGKRVIVACTOC---RQFKIRAKYKE-----DLESY 176
OY 59 LGGT-----GAFEIEINGOLVPSKLENGFPY 85
DB 177 VSGVSVIINPQKPRRGCLIEIREGGGVFISLNMRRPF 214

RESULT 14
US-10-510-386-36
; Sequence 36, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
```

APPLICANT: Jorgensen, Steen Troels
APPLICANT: Olsen, Peter Bjarke
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: Improved Bacillus Host Cell
FILE REFERENCE: 10294.204-US
CURRENT APPLICATION NUMBER: US/10/510,386
CURRENT FILING DATE: 2004-10-04
NUMBER OF SEQ ID NOS: 248
SOFTWARE: PatentIn version 3.3
SEQ ID NO 36
LENGTH: 329
TYPE: PRT
ORGANISM: Bacillus licheniformis
US-10-510-386-36

Query Match 11.5%; Score 68.5; DB 6; Length 329;
Best Local Similarity 20.5%; Pred. No. 8.3;
Matches 27; Conservative 22; Mismatches 36; Indels 47; Gaps 5;

QY 6 GQTSVAPPEEVEPGSVRIVEYCEPCGFATYIELASAVKEQYPGIEISRLGCTG-- 63
DB 168 GETAVITQPQOYNHGERTK-----GFEQT-----IKQKYPNMKVAAVLDGKDE 211
QY 64 -----AFBIEINGQL--VFSKLENG-----GPFYEKDLIEAIR 94
DB 212 LTSKKEAAKILIEENPSIKGIPTTEANGASGVAAVKEAGLEGEVCIIGFDKDKTLDDGK 271

QY 95 RASNGETLEKIT 106
DB 272 NGSISATMSQDT 283

RESULT 15
US-10-510-386-196
Sequence 196; Application US/10510386
Publication No. US20050244922A1
GENERAL INFORMATION:
APPLICANT: Andersen, Jens Tonne
APPLICANT: Clausen, Ib Groth
APPLICANT: Jorgensen, Steen Troels
APPLICANT: Olsen, Peter Bjarke
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: Improved Bacillus Host Cell
FILE REFERENCE: 10294.204-US
CURRENT APPLICATION NUMBER: US/10/510,386
CURRENT FILING DATE: 2004-10-04
NUMBER OF SEQ ID NOS: 248
SOFTWARE: PatentIn version 3.3
SEQ ID NO 196
LENGTH: 330
TYPE: PRT
ORGANISM: Bacillus licheniformis
US-10-510-386-196

Query Match 11.5%; Score 68.5; DB 6; Length 330;
Best Local Similarity 20.5%; Pred. No. 8.4;
Matches 27; Conservative 22; Mismatches 36; Indels 47; Gaps 5;

QY 6 GQTSVAPPEEVEPGSVRIVEYCEPCGFATYIELASAVKEQYPGIEISRLGCTG-- 63
DB 168 GETAVITQPQOYNHGERTK-----GFEQT-----IKQKYPNMKVAAVLDGKDE 211
QY 64 -----AFBIEINGQL--VFSKLENG-----GPFYEKDLIEAIR 94
DB 212 LTSKKEAAKILIEENPSIKGIPTTEANGASGVAAVKEAGLEGEVCIIGFDKDKTLDDGK 271

QY 95 RASNGETLEKIT 106
DB 272 NGSISATMSQDT 283

Search completed: March 27, 2006, 12:39:19
Job time : 13.7395 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2006, 11:54:20 ; Search time 119.82 Seconds
(without alignments)
421.703 Million cell updates/sec

Title: US-09-824-787b-2

Perfect score: 597
Sequence: 1 MSBEPGQTSVAPPEEVEPG.....ASNGETLEKITSRPPCVIL 115,

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_21: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *
9: geneseqp2005s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597	100.0	115	4	AAg78997 Human C35
2	597	100.0	115	4	AAg78780 Human C35
3	597	100.0	115	5	ABP58560 Human sit
4	597	100.0	115	5	ABP43843 RIKEN 181
5	597	100.0	115	8	ADH13244 Human mal
6	597	100.0	115	8	ADK48916 Human bre
7	597	100.0	115	9	ADK48916 Human bre
8	597	100.0	115	9	ADK48916 Human bre
9	597	100.0	115	9	ADK48916 Human bre
10	597	100.0	115	9	ADK48916 Human bre
11	597	100.0	115	9	ADK48916 Human bre
12	597	100.0	115	9	ADK48916 Human bre
13	597	100.0	115	9	ADK48916 Human bre
14	597	100.0	115	9	ADK48916 Human bre
15	597	100.0	115	9	ADK48916 Human bre
16	597	100.0	115	9	ADK48916 Human bre
17	597	100.0	115	9	ADK48916 Human bre
18	597	100.0	115	9	ADK48916 Human bre
19	597	100.0	115	9	ADK48916 Human bre
20	597	100.0	115	9	ADK48916 Human bre
21	597	100.0	115	9	ADK48916 Human bre
22	597	100.0	115	9	ADK48916 Human bre
23	597	100.0	115	9	ADK48916 Human bre
24	597	100.0	115	9	ADK48916 Human bre

25	289	48.4	73	8	ADK50214 Human car
26	289	48.4	73	8	ADK50202 Human car
27	288	48.2	64	8	ADK50187 Human car
28	288	48.2	64	8	ADK50213 Human car
29	288	48.2	73	8	ADK50188 Human car
30	288	48.2	75	8	ADK50196 Human car
31	287	48.1	64	8	ADK50209 Human car
32	287	48.1	64	8	ADK50215 Human car
33	287	48.1	73	8	ADK50210 Human car
34	287	48.1	73	8	ADK50216 Human car
35	286	47.9	64	8	ADK50227 Human car
36	286	47.9	64	8	ADK50221 Human car
37	286	47.9	66	8	ADK50225 Human car
38	286	47.9	73	8	ADK50222 Human car
39	286	47.9	73	8	ADK50182 Human car
40	286	47.9	73	8	ADK50228 Human car
41	286	47.9	77	8	ADK50226 Human car
42	285.5	47.8	85	8	ADK50206 Human car
43	285	47.7	64	8	ADK50229 Human car
44	285	47.7	65	8	ADK50189 Human car
45	285	47.7	73	8	ADK50230 Human car

ALIGNMENTS

RESULT 1
ID AAg78997 standard; protein; 115 AA.
AC AAg78997;
XX
DT 22-JUN-2002 (first entry)
DE Human C35, a tumour antigen.
XX
KM Human; C35; cytosolic; gene therapy; vaccine; tumour antigen;
KM breast cancer; bladder cancer; tumour immunotherapy; chromosome 17q12.
XX
OS Homo sapiens.
XX
PN WO200174859-A2.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-US010855.
XX
PR 04-APR-2000; 2000US-0194463P.
XX
PA (UYRP) UNIV ROCHESTER.
XX
PI Zauderer M, Evans EE, Borrello MA;
XX
DR WPI; 2001-626383/72.
XX
DR N-PSDB; AA171785, AA171793.
PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions
PT and vaccines, for inducing antibody and cell-mediated immunity against
PT target cells, such as tumor cells that express C35 gene.
XX
PS Claim 11; Fig 1; 33pp; English.
XX
CC The present sequence is human C35. C35 is a novel tumour antigen that is
CC overexpressed in human breast and bladder carcinoma. C35 is thought to be
CC a promising candidate for tumour immunotherapy, in immunogenic
CC compositions and vaccines, to induce antibody and cell-mediated immunity
CC against target cells such as tumour cells that express C35 genes. The C35
CC gene aligns on human chromosome 17q12
XX
SQ Sequence 115 AA;
XX
Query Match 100.0%; Score 597; DB 4; Length 115;
Best local Similarity 100.0%; Pred. No. 1.3e-60;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGEPGQTSVAPPPEVEPGSGVRIVVEYCEPGFATYIELASAVKEOYPGIIESRLG 60
DB 1 MSGEPGQTSVAPPPEVEPGSGVRIVVEYCEPGFATYIELASAVKEOYPGIIESRLG 60

OY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGTELEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGTELEKITSRPPCVIL 115

RESULT 2
AAG7870
ID AAG7870 standard; protein; 115 AA.
AC AAG7870;
DT 08-MAY-2002 (first entry)
DE Human C35 protein.

XX C35 protein; antigenic peptide; major histocompatibility complex;
KW MHC-peptide complex; MHC; human; MHC class I alpha chain;
KW beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;
KW vaccine; immune response modulation; hyperproliferative disorder;
KW neoplasm; hypergammaglobulinemia; viral infection; hepatitis;
KW meningitis; bacterial infection; tuberculosis; gingivitis;
KW parasitic infection; autoimmune disease; Hashimoto's disease;
KW Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;
KW graft-versus-host disease; GVHD; breast cancer.

OS Homo sapiens.
XX
XX W0200178768-A2.
XX
XX 25-OCT-2001.
XX
XX 12-APR-2001; 2001WO-US011912.
XX
XX 12-APR-2000; 2000US-0196472P.
XX
XX (UVRP) UNIV ROCHESTER.
XX
XX Zauderer M, Smith ES;
XX
XX WPI; 2001-602927/68.
XX
XX N-PSDB; AAH77148.
XX
XX Novel compound comprising major histocompatibility complex-peptide
XX PT complexes, used to modulate immune responses.
XX
XX Example 16; Fig 7; 16pp; English.

XX The invention comprises a compound which contains one or more major
XX CC histocompatibility complex (MHC)-peptide complexes, and an antibody
XX CC specific for a cell surface marker. The complexes comprise an MHC class I
XX CC alpha chain, a beta-2 microglobulin molecule and an antigenic peptide
XX CC bound in the MHC groove. Alternatively, the complexes may comprise an MHC
XX CC class II alpha chain, an MHC class II beta chain, and an antigenic
XX CC peptide bound in the MHC groove. The complexes are linked to the carboxyl
XX CC terminus of the antibody. The compounds of the invention can be used as a
XX CC vaccine to modulate an immune response. The compounds of the invention
XX CC are useful for treating: hyperproliferative disorders (e.g. neoplasms and
XX CC hypergammaglobulinemia); viral infections (e.g. hepatitis and
XX CC meningitis); bacterial infections (e.g. tuberculosis and gingivitis);
XX CC parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,
XX CC Graves' disease and rheumatoid arthritis); allergic reactions/conditions
XX CC (e.g. asthma). The compounds of the invention may also be used in the
XX CC treatment of organ rejection or graft-versus-host disease (GVHD). The
XX CC present sequence represents C35 protein, this protein is differentially
XX CC expressed in human breast cancer cells. In one embodiment of the
XX CC invention, antigenic peptides derived from the C35 protein can be used to
XX CC promote an immune response against a cancerous cell.

XX Sequence 115 AA;
SQ

Query Match 100.0%; Score 597; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.3e-60;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSGEPGQTSVAPPPEVEPGSGVRIVVEYCEPGFATYIELASAVKEOYPGIIESRLG 60
DB 1 MSGEPGQTSVAPPPEVEPGSGVRIVVEYCEPGFATYIELASAVKEOYPGIIESRLG 60

OY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGTELEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGTELEKITSRPPCVIL 115

RESULT 3
ABP58560
ID ABP58560 standard; protein; 115 AA.
AC ABP58560;
DT 28-MAR-2003 (first entry)
DE Human site-specific recombinase motif-containing protein 12.63.

XX Human, site-specific recombinase motif-containing protein 12.63;
KW recombinant production; gene therapy; cancer; tumour; HIV infection;
KW human immunodeficiency virus; cytostatic.

OS Homo sapiens.
XX
XX CN1363589-A.
XX
XX 14-AUG-2002.
XX
XX 05-JAN-2001; 2001CN-00105072.
XX
XX 05-JAN-2001; 2001CN-00105072.
XX
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2002-742038/81.
XX
XX N-PSDB; ABZ25755.
XX
XX Polypeptide-human protein 12.63 containing site-specific recombinase
XX PT characteristic sequence fragment and polynucleotide for coding it.
XX
XX Claim 1, Page 27 (Disclosure); 33pp; Chinese.

XX The invention relates to human site-specific recombinase motif-containing
XX CC protein 12.63 (ABP58560) and nucleic acids encoding it (ABZ25755). The
XX CC protein has a molecular weight of 12.63 kD. The invention also relates to
XX CC a method for the recombinant production of the protein, an antagonist of
XX CC the protein, and the use of the protein, gene and antagonist in
XX CC therapeutic applications. Site-specific recombinase motif-containing
XX CC protein 12.63 can be used in the treatment of a variety of diseases such
XX CC as cancer and HIV (human immunodeficiency virus) infection. The present
XX CC sequence represents human site-specific recombinase motif-containing
XX CC protein 12.63

SQ Sequence 115 AA;
SQ

Query Match 100.0%; Score 597; DB 5; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.3e-60;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSGEPGQTSVAPPPEVEPGSGVRIVVEYCEPGFATYIELASAVKEOYPGIIESRLG 60
DB 1 MSGEPGQTSVAPPPEVEPGSGVRIVVEYCEPGFATYIELASAVKEOYPGIIESRLG 60

QY 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115
DB 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115

RESULT 4

ID ABP43843 standard; protein, 115 AA.

AC ABP43843;

DT 26-FEB-2003 (first entry)

DE RIKEN 1810046J19 protein.

XX Neuroprotective; immunomodulator; cancer; chromosome 4q13-q21;
XX cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
XX wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
XX amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
XX vlnetary.

OS Homo sapiens.

PN W0200231111-A2.

PD 18-APR-2002.

PF 11-OCT-2001; 2001MO-US027760.

PR 12-OCT-2000; 2000US-00687527.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Dymnac RT;

DR WPI; 2002-426278/45.

XX N-PSDB; AB061067.

PT New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.

PS Claim 20; SEQ ID # 746; 357pp + Sequence listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The
XX activity of polynucleotides of the invention may be described as,
XX vulnetary, neuroprotective, immunomodulator, cyostatic and anti-
XX inflammatory. Compositions comprising nucleic acids of the invention are
XX useful for treating a mammalian subject, or as nutritional sources or
XX supplements. These are useful in gene therapy, particularly for treating
XX wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, autoimmune disorders, cancer or
XX inflammation. The nucleic acids and polypeptides are also useful in
XX diagnostic and research methods. The sequences given in records ABP43544-
XX ABP43989 represent polypeptides encoded by polynucleotides of the
XX invention. NOTE: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WPIO at fcp.wipo.int/pub/published_pct_sequences

XX Sequence 115 AA;

Query Match 100.0%; Score 597; DB 5; Length 115;

Best Local Similarity 100.0%; Pred. No. 1.3e-60; Indels 0; Gaps 0;
Matches 115; Conservative 0; Mismatches 0;

QY 1 MSGEPQOTSVAAPPEVEBPGSVRIWVCEPCGFATYIELASAVKEQYPGIEISRIG 60
DB 1 MSGEPQOTSVAAPPEVEBPGSVRIWVCEPCGFATYIELASAVKEQYPGIEISRIG 60
QY 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115
DB 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115

DB 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115

RESULT 5

ID ADH13244 standard; protein, 115 AA.

AC ADH13244;

DT 11-MAR-2004 (first entry)

DE Human malignant neoplasia-related protein SegID93.

XX malignant neoplasia; cyostatic; breast cancer; ovarian cancer;
XX gastric cancer; colon cancer; oesophageal cancer; mesenchymal cancer;
XX bladder cancer; non-small cell lung cancer; human.

OS Homo sapiens.

PN EPI365034-A2.

PD 26-NOV-2003.

PF 09-MAY-2003; 2003EP-00010447.

PR 21-MAY-2002; 2002EP-00010291.

PR 13-FEB-2003; 2003EP-00003112.

PA (FARB) BAYER AG.

PI Wirtz R, Munnes M, Kallabis H;

DR WPI; 2004-073279/08.

XX N-PSDB; ADH13221.

PT Predicting, diagnosing or prognosing malignant neoplasia by detecting at
PT least two markers, where the markers are genes from one or more
PT chromosomal regions altered in malignant neoplasia..

PS Claim 11; SEQ ID NO 93; 267pp; English.

XX This invention relates to a novel method for the prediction, diagnosis,
XX or prognosis of malignant neoplasia by the detection of at least two
XX markers. The invention may also be useful for the development of
XX cyostatic compounds through the regulation of the expression of a gene
XX or activity of a protein associated with malignant neoplasia. The method
XX is useful for prediction, diagnosis or prognosis of malignant neoplasia
XX such as breast cancer, ovarian cancer, gastric cancer, colon cancer,
XX oesophageal cancer, mesenchymal cancer, bladder cancer or non-small cell
XX lung cancer. The polynucleotides and polypeptides defined in the
XX specification, antisense polynucleotides targeting the polynucleotides,
XX antibodies targeting either one of the polynucleotides or polypeptides,
XX and compounds identified by the screening methods are useful for
XX preventing or treating malignant neoplasia. The disease treated is
XX preferably breast cancer. The present sequence is that of a human
XX malignant neoplasia-related protein which may be used in the method of
XX the invention.

XX Sequence 115 AA;

Query Match 100.0%; Score 597; DB 8; Length 115;

Best Local Similarity 100.0%; Pred. No. 1.3e-60; Indels 0; Gaps 0;
Matches 115; Conservative 0; Mismatches 0;

QY 1 MSGEPQOTSVAAPPEVEBPGSVRIWVCEPCGFATYIELASAVKEQYPGIEISRIG 60
DB 1 MSGEPQOTSVAAPPEVEBPGSVRIWVCEPCGFATYIELASAVKEQYPGIEISRIG 60
QY 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115
DB 61 GTGAFFIEINGQLVFSKLENGGFPYKDLIEAIRRASNGETLEKITSRPPCVIL 115

RESULT 6
ADK48916
ID ADK48916 standard; protein; 115 AA.
XX
XX ADK48916;
XX
XX 04-NOV-2004 (first entry)
DT
XX Human breast/bladder carcinoma-related C35 protein.
DE
XX C35 epitope; cytostatic; vaccine; tumour; breast; bladder carcinoma;
KM human; C35.
XX
XX Homo sapiens.
OS
XX MO2003104428-A2.
PN
XX 18-DEC-2003.
PD
XX 10-JUN-2003; 2003WO-US018252.
PF
XX 10-JUN-2002; 2002US-0386738P.
PR 11-DEC-2002; 2002US-0432241P.
PR 23-APR-2003; 2003US-0464650P.
XX
XX (VACC-) VACCINEX INC.
PA (VYRP) UNITV ROCHESTER.
XX
XX Zauderer M, Evans EE, Borrello MA;
PI WPI; 2004-062349/06.
DR N-PSDB; ADK48915.
XX
XX Novel C35 polypeptide useful for formulation of immunogenic composition
PT to induce antibodies and cell-mediated immunity against tumor cells.
XX
XX Claim 1; SEQ ID NO 2; 626pp; English.
PS
XX The invention relates to a novel isolated polypeptide comprising or
CC consisting of two or more C35 peptide epitopes. The polypeptide of the
CC invention demonstrates cytostatic activity and may be useful for the
CC formulation of an immunogenic composition, such as a vaccine, to induce
CC antibodies and cell-mediated immunity against target cells such as tumor
CC cells. Furthermore, the polypeptide and its analogues may be useful as
CC prognostic markers for carcinoma, such as human breast or bladder
CC carcinoma. The current sequence is that of human breast/bladder carcinoma
CC -related C35 protein of the invention.
XX
XX Sequence 115 AA;
SQ
Query Match 100.0%; Score 597; DB 8; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.3e-60;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSGEPGQTSVAPPEEVEPGSGVRIIVEYCEPGFEATYLELSAVKEQYPGIEISRLG 60
DB 1 MSGEPPGQTSVAPPEEVEPGSGVRIIVEYCEPGFEATYLELSAVKEQYPGIEISRLG 60
QY 61 GTGAFFIEINGQVFSKLENGGFPYRKDLIEAIRRANGETLEKITSRPPCVIL 115
DB 61 GTGAFFIEINGQVFSKLENGGFPYRKDLIEAIRRANGETLEKITSRPPCVIL 115
RESULT 7
ADV60511
ID ADV60511 standard; protein; 115 AA.
XX
XX ADV60511;
XX
XX 24-FEB-2005 (first entry)
DT
XX Human breast cancer staging marker protein Seq 86.
DE
XX

KM breast tumor; cytostatic; diagnosis; prognosis; tumour marker; cancer.
XX
XX Homo sapiens.
OS
XX MO2004106495-A2.
PN
XX 09-DEC-2004.
PD
XX 26-MAY-2004; 2004WO-US016793.
PF
XX 29-MAY-2003; 2003US-0474281P.
PR 23-MAR-2004; 2004US-0555557P.
XX
XX (MIL-) MILLENNIUM PHARM INC.
PA
XX Mohanan JE, Hoersch S, Anderson DL, Endege WO, Ford D, Giatt K;
PI Gorbarcheva BO, Kamatkar S, Xu Y, Gannavaram M, Zhao X, Schlegel R;
PI Mertens M;
XX
XX WPI; 2005-039718/04.
DR N-PSDB; ADV60510.
XX
XX Novel marker protein, useful for assessing whether patient is afflicted
PT with breast cancer, for assessing efficacy of therapy for inhibiting
PT breast cancer, for assessing breast cell carcinogenic potential of test
PT composition.
XX
XX Disclosure; SEQ ID NO 86; 279pp; English.
PS
XX This invention relates to a novel secreted human marker proteins (Mks)
CC and the encoding nucleic acid molecules thereof. Specifically, it refers
CC to a method for assessing whether a patient has breast cancer that has
CC metastasized or is likely to metastasize. The present invention describes
CC determining the level of expression of a marker protein in the patient
CC sample and comparing this to the level from a control subject having a
CC non-metastasized breast tumor or no breast tumor. Furthermore it provides
CC a screening method for assessing the suitability of one or more test
CC compounds at inhibiting breast cancer in a patient by inhibiting the
CC expression of the marker proteins as given in the specification. As such,
CC it is useful for predicting the clinical outcome of a breast cancer
CC patient, for monitoring progression of the disease and for assessing the
CC breast cell carcinogenic potential of a test composition. Accordingly,
CC pharmaceutical compositions derived thereof exhibit cytostatic
CC activities. This polypeptide is a human breast cancer staging marker
CC protein of the invention.
XX
XX Sequence 115 AA;
SQ
Query Match 100.0%; Score 597; DB 9; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.3e-60;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSGEPGQTSVAPPEEVEPGSGVRIIVEYCEPGFEATYLELSAVKEQYPGIEISRLG 60
DB 1 MSGEPPGQTSVAPPEEVEPGSGVRIIVEYCEPGFEATYLELSAVKEQYPGIEISRLG 60
QY 61 GTGAFFIEINGQVFSKLENGGFPYRKDLIEAIRRANGETLEKITSRPPCVIL 115
DB 61 GTGAFFIEINGQVFSKLENGGFPYRKDLIEAIRRANGETLEKITSRPPCVIL 115
RESULT 8
ADX83703
ID ADX83703 standard; protein; 115 AA.
XX
XX ADX83703;
XX
XX 05-MAY-2005 (first entry)
DT
XX Human C35 antigenic protein, seqid:2.
DE
XX Vaccine; immune modulation; delivery mechanism; microglobulin;
KM chemotherapy; gene therapy; adoptive immunotherapy;
XX

KM major histocompatibility complex; cancer; cytostatic; neoplasm;
 KM infectious disease; antimicrobial; infection; autoimmune disease;
 XX immunosuppressive; immune disorder; allergy; antiallergic; C35 DNA.
 OS Homo sapiens.
 XX US2005042218-A1.
 PN 24-FEB-2005.
 PD 09-JUL-2004; 2004US-00887230.
 PF 10-JUL-2003; 2003US-0485746P.
 PR 22-OCT-2003; 2003US-0513043P.
 XX (VACC-) VACCINEX INC.
 PA Zauderer M;
 PI WPI; 2005-180769/19.
 DR N-PSDB; ADX83702.
 XX New compounds comprising major histocompatibility complex Class I-peptide
 PT -antibody conjugates with modified beta2-microglobulin, useful for
 PT modulating immune responses or for treating or preventing e.g. cancer or
 PT infections.
 XX Disclosure; SEQ ID NO 2; 65pp; English.
 PS The present invention provides a novel targeted vaccine delivery system
 CC comprising one or more peptide-major histocompatibility complex (MHC)
 CC Class I complexes linked through the beta2-microglobulin molecule to an
 CC antibody which is specific for a cell surface marker. The invention is
 CC useful for modulating an immune response and for preventing and treating
 CC cancer, infectious diseases, autoimmune diseases and allergies. The
 CC present sequence is human C35 antigenic protein.
 XX Sequence 115 AA;
 SQ

Query Match 100.0%; Score 597; DB 9; Length 115;
 Best Local Similarity 100.0%; Pred. No. 1.3e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGEPGQTSVAPPEVEPESGVRIVVEYCEPGFATYLELASAVKEQYPGIEISRLG 60
 DB 1 MSGEPGQTSVAPPEVEPESGVRIVVEYCEPGFATYLELASAVKEQYPGIEISRLG 60
 QY 61 GTGAFIEIINGQLVFSKLENGGFPYKDLIEAIRASNGETLEKITSRPPCVIL 115
 DB 61 GTGAFIEIINGQLVFSKLENGGFPYKDLIEAIRASNGETLEKITSRPPCVIL 115

RESULT 9
 AEA15131
 ID AEA15131 standard; protein; 115 AA.
 AC AEA15131;
 XX 28-JUL-2005 (first entry)
 DT Human polypeptide #44.
 DE Human polypeptide #44.
 KM Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;
 KM colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;
 KM cytostatic; neoplasm.
 XX Homo sapiens.
 OS Homo sapiens.
 XX MO2005047534-A2.
 PN 26-MAY-2005.
 PD 15-OCT-2004; 2004WO-BP011599.
 PF

XX 28-OCT-2003; 2003EP-00024565.
 PR (FARB) BAYER HEALTHCARE AG.
 PA Wirtz R, Munnes M;
 XX WPI; 2005-372393/38.
 DR N-PSDB; AEA15108.
 XX Predicting a response to cancer treatment by detecting at least 2
 PT markers, which are genes or genomic nucleic acid sequences that are
 PT located on one chromosomal region, which is altered in malignant
 PT neoplasia.
 XX Claim 8; SEQ ID NO 93; 464pp; English.
 PS The invention relates to a method of predicting response to cancer
 CC treatment comprising detection of at least 2 markers, where the markers
 CC are genes and fragments or genomic nucleic acid sequences that are
 CC located on one chromosomal region, which is altered in malignant
 CC neoplasia. The invention also relates to a method for the prediction,
 CC diagnosis or prognosis of malignant neoplasia, methods for detecting
 CC deregulations in malignant neoplasia and breast cancer, a method of
 CC determining the phenotype of a cell or tissue, a method for identifying
 CC genomic regions which are altered on the chromosomal level and encode
 CC genes that are linked by function and are differentially expressed in
 CC malignant neoplasia and breast cancer, methods of screening for agents
 CC which regulate the activity of a polypeptide or a polynucleotide and
 CC antibodies that specifically bind to a full length or partial
 CC polypeptide. The method is useful for predicting response to cancer
 CC treatment. The method and compositions are useful for predicting,
 CC diagnosing, prognosing, preventing or treating malignant neoplasia
 CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,
 CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell
 CC lung cancer. This sequence represents a human polypeptide used in the
 CC scope of the invention.
 XX Sequence 115 AA;
 SQ

Query Match 100.0%; Score 597; DB 9; Length 115;
 Best Local Similarity 100.0%; Pred. No. 1.3e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGEPGQTSVAPPEVEPESGVRIVVEYCEPGFATYLELASAVKEQYPGIEISRLG 60
 DB 1 MSGEPGQTSVAPPEVEPESGVRIVVEYCEPGFATYLELASAVKEQYPGIEISRLG 60
 QY 61 GTGAFIEIINGQLVFSKLENGGFPYKDLIEAIRASNGETLEKITSRPPCVIL 115
 DB 61 GTGAFIEIINGQLVFSKLENGGFPYKDLIEAIRASNGETLEKITSRPPCVIL 115

RESULT 10
 AEA62447
 ID AEA62447 standard; protein; 115 AA.
 AC AEA62447;
 XX 25-AUG-2005 (first entry)
 DT Human C35 antigen.
 DE Human C35 antigen.
 KM C35 antigen; antigen; cell growth; cancer; cytostatic; apoptosis;
 KM immunotherapy; hyperproliferation.
 XX Homo sapiens.
 OS Homo sapiens.
 XX MO2005055936-A2.
 PN 23-JUN-2005.
 PD 06-DEC-2004; 2004WO-US040573.
 PF

XX 04-DEC-2003; 2003US-0526572P.
 PR 23-DEC-2003; 2003US-0531688P.
 XX (VACC-) VACCIMEX INC.
 PA Evans EE, Parle MJ, Sahaarabudhe DM, Smith ES, Zauderer M;
 PI WPI; 2005-458501/46.
 DR N-PSDB; ABA62446.
 XX
 XX Killing cancer cells, by administering apoptosis-inducing therapy and
 PT administering antibody specific for intracellular, cancer-associated
 PT protein other than C35, or antibody specific for C35.
 XX
 PS Disclosure; SEQ ID NO 2; 255pp; English.
 XX
 CC The invention relates to killing (M1) cancer cells, comprising
 CC administering an apoptosis-inducing therapy to cancer cells, and
 CC administering to the cells an antibody specific for an intracellular,
 CC cancer-associated protein, provided that the protein is not C35 antigen,
 CC where protein becomes exposed on the cell surface in cells undergoing
 CC apoptosis, where the antibody is conjugated to or complexed with a toxin.
 CC The non-C35 antigen protein is a prenylated protein. Also included are an
 CC isolated antibody (1) specific for C35 (chosen from an antibody
 CC comprising the VH region encoded by clone 1B3G, the VL region encoded by
 CC clone 1B3K, the VH region encoded by clone 1B3G, the VL region encoded by
 CC clone 1F2K, the VH region encoded by clone H0009, the VL region encoded
 CC by clone L0010, an antibody comprising at least one of CDR1 or CDR2 of
 CC the VH region encoded by ABA62495, an antibody comprising at least one of
 CC CDR1 or CDR2 of the VH region encoded by ABA62499, an antibody comprising
 CC at least one of CDR1, CDR2, or CDR3 of the VL region encoded by
 CC ABA62497, a chimeric antibody, or a humanized antibody), a polynucleotide
 CC encoding the antibody, a vector comprising the polynucleotide, a host
 CC cell comprising the vector and a composition comprising the antibody and
 CC a carrier. The method is useful for killing cancer cells in a mammal
 CC preferably human in need of eradication of smaller tumors and/or
 CC micrometastases, or in need of cancer treatment for C35-associated cancer
 CC chosen from breast cancer, ovarian cancer, bladder cancer, lung cancer,
 CC prostate cancer, pancreatic cancer, colon cancer, melanoma and other
 CC hyperproliferative disorders. The antibody is useful for detecting,
 CC diagnosing or monitoring C35-associated cancer. The antibody comprises a
 CC chimeric antibody comprising human immunoglobulin constant regions fused
 CC to the variable regions of mouse anti-35 antibodies (named 1F2, 1B3, Mab
 CC 165 and Mab 171). The present sequence represents the human C35 antigen.
 XX
 XX Sequence 115 AA;
 S0
 Query Match 100.0%; Score 597; DB 9; Length 115;
 Best Local Similarity 100.0%; Pred. No. 1,3e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSGPGGTGVAAPPEEVEPGSGVRIVVEYCEPCGFATYIELASVKEQYFGIIEIRLG 60
 DB 1 MSGPGGTGVAAPPEEVEPGSGVRIVVEYCEPCGFATYIELASVKEQYFGIIEIRLG 60
 QY 61 GTGAFELIINGQLVFSKLENGGFPYKDLIAIRASNGELTEKTRSPCVLL 115
 DB 61 GTGAFELIINGQLVFSKLENGGFPYKDLIAIRASNGELTEKTRSPCVLL 115
 RESULT 11
 ABP43055
 ID ABP43055 standard; protein; 124 AA.
 XX
 AC ABP43055;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HVCAA68, SEQ ID NO:4187.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX
 OS Homo sapiens.
 XX
 XX WO200200677-A1.
 XX
 XX 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US018569.
 XX
 XX 07-JUN-2000; 2000US-0209467P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Birse CE, Rosen CA;
 XX
 XX WPI; 2002-147878/19.
 DR N-PSDB; ABO56132.
 XX
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX
 PS Claim 11; SEQ ID NO 4187; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 S0
 Query Match 100.0%; Score 597; DB 5; Length 124;
 Best Local Similarity 100.0%; Pred. No. 1.4e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSGPGGTGVAAPPEEVEPGSGVRIVVEYCEPCGFATYIELASVKEQYFGIIEIRLG 60
 DB 10 MSGPGGTGVAAPPEEVEPGSGVRIVVEYCEPCGFATYIELASVKEQYFGIIEIRLG 69

Oy 61 GTGAFEIINGQLVFSKLENGGFPPEKDLIEAIRRANGETLEKITSRPPCVIL 115
 DB 70 GTGAFEIINGQLVFSKLENGGFPPEKDLIEAIRRANGETLEKITSRPPCVIL 124

RESULT 12

AAB43521
 ID AAB43521 standard; protein; 131 AA.

XX AAB43521;

XX 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO:966.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
 KM diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KM antidiabetic; antiaesthetic; antirheumatic; antithyroid; antiviral;
 KM antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;
 KM dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 KM vasotropic; antipruritic; antidiabetic; gene therapy; inflammation;
 KM allergic reaction; graft versus host disease; organ rejection;
 KM hemostatic; thrombolytic; cardiovascular disorder; infection;
 KM neurological disease; drug screening.

XX Homo sapiens.

XX W020005350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005882.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-58753/55.

XX N-PSDB; AAC77730.

XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer.

XX Claim 11; Page 1534-1535; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given in
 CC AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnery; immunomodulator;
 CC antidiabetic; antiaesthetic; antirheumatic; antithyroid; antiviral;
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;
 CC dermatological; neuroprotective; thrombolytic; coagulant;
 CC neotropic; vasotropic; antipruritic; antidiabetic; gene therapy; inflammation;
 CC allergic reaction; graft versus host disease; organ rejection;
 CC hemostatic; thrombolytic; cardiovascular disorder; infection;
 CC neurological disease; drug screening.
 CC The present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44440 represent sequences used in the exemplification of
 CC the present invention

XX Sequence 131 AA;

XX Query Match 100.0%; Score 597; DB 3; Length 131;

Best Local Similarity 100.0%; Pred. No. 1.5e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSGEPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYLELASAVKEQYIGIIESRLG 60

DB 17 MSGEPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYLELASAVKEQYIGIIESRLG 76

Oy 61 GTGAFEIINGQLVFSKLENGGFPPEKDLIEAIRRANGETLEKITSRPPCVIL 115
 DB 77 GTGAFEIINGQLVFSKLENGGFPPEKDLIEAIRRANGETLEKITSRPPCVIL 131

RESULT 13

ID ADK49069 standard; protein; 131 AA.

XX ADK49069;

XX 04-NOV-2004 (first entry)

XX Human breast/bladder carcinoma C35-related protein - SEQ ID 155.

XX C35 epitope; cytostatic; vaccine; tumour; breast; bladder carcinoma;

XX human.

XX Homo sapiens.

XX W02003104428-A2.

XX 18-DEC-2003.

XX 10-JUN-2003; 2003WO-US018252.

XX 10-JUN-2002; 2002US-0386738P.

XX 11-DEC-2002; 2002US-0432241P.

XX 23-APR-2003; 2003US-0464650P.

XX (VACC-) VACCINE INC.

XX (VIRP) UNIT ROCHESSTER.

XX Zauderer M, Evans EE, Borrello MA;

XX WPI; 2004-062349/06.

XX Novel C35 polypeptide useful for formulation of immunogenic composition

PT to induce antibodies and cell-mediated immunity against tumor cells.

XX Disclosure; SEQ ID NO 155; 626pp; English.

XX The invention relates to a novel isolated polypeptide comprising or
 CC consisting of two or more C35 peptide epitopes. The polypeptide of the
 CC invention demonstrates cytostatic activity and may be useful for the
 CC formulation of an immunogenic composition, such as a vaccine, to induce
 CC antibodies and cell-mediated immunity against target cells such as tumour
 CC cells. Furthermore, the polypeptide and its analogues may be useful as
 CC prognostic markers for carcinoma, such as human breast or bladder
 CC carcinoma. The current sequence is that of human breast/bladder carcinoma
 CC C35-related protein of the invention.

XX Sequence 131 AA;

Query Match 100.0%; Score 597; DB 8; Length 131;
 Best Local Similarity 100.0%; Pred. No. 1.5e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSGEPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYLELASAVKEQYIGIIESRLG 60

DB 17 MSGEPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYLELASAVKEQYIGIIESRLG 76

Oy 61 GTGAFEIINGQLVFSKLENGGFPPEKDLIEAIRRANGETLEKITSRPPCVIL 115
 DB 77 GTGAFEIINGQLVFSKLENGGFPPEKDLIEAIRRANGETLEKITSRPPCVIL 131

RESULT 14
 AEA62539
 ID AEA62539 standard; protein; 149 AA.
 XX
 AC AEA62539;
 XX
 DT 25-AUG-2005 (first entry)
 XX
 DE Recombinant C35 antigen.
 XX
 KW C35 antigen; antigen; cell growth; cancer; cytostatic; apoptosis;
 KM immunotherapy; hyperproliferation; epitope mapping.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..29
 FT /note= "His-tagged signal peptide"
 FT Protein 30..149
 FT /note= "C35 antigen"
 XX
 FN WO2005055936-A2.
 XX
 PD 23-JUN-2005.
 XX
 PF 06-DEC-2004; 2004WO-US040573.
 XX
 PR 04-DEC-2003; 2003US-0526572P.
 PR 23-DEC-2003; 2003US-0531688P.
 XX
 PA (VACC-) VACCINEX INC.
 XX
 PI Evans EE, Paris MJ, Saharabudhe DM, Smith ES, Zauderer M;
 XX
 DR WPI; 2005-458501/46.
 XX
 PT Killing cancer cells, by administering apoptosis-inducing therapy and
 PT administering antibody specific for intracellular, cancer-associated
 PT protein other than C35, or antibody specific for C35.
 XX
 PS Example 11; Fig 9; 255pp; English.
 XX
 CC The invention relates to killing (M1) cancer cells, comprising
 CC administering an apoptosis-inducing therapy to cancer cells, and
 CC administering to the cells an antibody specific for an intracellular,
 CC cancer-associated protein, provided that the protein is not C35 antigen,
 CC where protein becomes exposed on the cell surface in cells undergoing
 CC apoptosis, where the antibody is conjugated to or complexed with a toxin.
 CC The non-C35 antigen protein is a prenylated protein. Also included are an
 CC isolated antibody (1) specific for C35 (chosen from an antibody
 CC comprising the VH region encoded by clone 1B3G, the VL region encoded by
 CC clone 1B3K, the VH region encoded by clone 1F2G, the VL region encoded by
 CC clone 1F2K, the VH region encoded by clone H0009, the VL region encoded
 CC by clone L0010, an antibody comprising at least one of CDR1 or CDR2 of
 CC the VH region encoded by AEA62495, an antibody comprising at least one of
 CC CDR1 or CDR2 of the VH region encoded by AEA62499, an antibody comprising
 CC at least one of CDR1, CDR2, or CDR3 of the VL region encoded by
 CC AEA62497, a chimeric antibody, or a humanized antibody), a polynucleotide
 CC encoding the antibody, a vector comprising the polynucleotide, a host
 CC cell comprising the vector and a composition comprising the antibody and
 CC a carrier. The method is useful for killing cancer cells in a mammal
 CC preferably human in need of eradication of smaller tumors and/or
 CC micrometastases, or in need of cancer treatment for C35-associated cancer
 CC chondrosarcoma, breast cancer, ovarian cancer, bladder cancer, lung cancer,
 CC prostate cancer, pancreatic cancer, colon cancer, melanoma and other
 CC hyperproliferative disorders. The antibody is useful for detecting,
 CC diagnosing or monitoring C35-associated cancers. The antibody comprises a
 CC chimeric antibody comprising human immunoglobulin constant regions fused
 CC to the variable regions of mouse anti-35 antibodies (named 1F2, 1B3, MAB
 CC 165 and MAb 171). The present sequence represents a recombinant human C35
 CC antigen (with a used His-tagged signal peptide), Lys-C digests of which

CC were used in epitope mapping studies.
 XX
 SQ Sequence 149 AA:
 XX
 Query Match 100.0%; Score 597; DB 9; Length 149;
 Best Local Similarity 100.0%; Pred. No. 1,8e-60;
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 MSGEPGQTSVAPPEEVEPGSGVRIIVEYCEPGCFEATYLEASAVKQYGIETESRLG 60
 DB 35 MSGEPGQTSVAPPEEVEPGSGVRIIVEYCEPGCFEATYLEASAVKQYGIETESRLG 94
 QY 61 GTGAFEEIINGQLVFSKLENGGPFPEYKDLLEAIRRANGETLEKRTNSRPPCVLL 115
 DB 95 GTGAFEEIINGQLVFSKLENGGPFPEYKDLLEAIRRANGETLEKRTNSRPPCVLL 149
 XX
 RESULT 15
 ABR47619
 ID ABR47619 standard; protein; 206 AA.
 XX
 AC ABR47619;
 XX
 DT 12-JUN-2003 (first entry)
 XX
 DE Breast cancer associated protein sequence SEQ ID NO:480.
 XX
 KM Human; breast cancer; cytostatic; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO2003004989-A2.
 XX
 PD 16-JAN-2003.
 XX
 PF 21-JUN-2002; 2002WO-US019669.
 XX
 PR 21-JUN-2001; 2001US-0299887P.
 PR 27-JUN-2001; 2001US-0301572P.
 PR 18-JUL-2001; 2001US-0306501P.
 PR 25-SEP-2001; 2001US-0325002P.
 PR 05-MAR-2002; 2002US-0362585P.
 PR 14-MAY-2002; 2002US-0380391P.
 XX
 PA (MILL-) MILLENIUM PHARM INC.
 XX
 PI Lillie J, Gannavarapu M, Glatt K, Hoerth S, Kamackar S;
 PI Mertens M, Monahan JR, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
 PI Baet RC, Hortobagyi GN, Puzetel L, Meric F, Sahin A, Mills GB;
 XX
 DR WPI; 2003-210381/20.
 DR N-FSDB; ACC50321.
 XX
 PT Breast cancer diagnosis or treatment by comparing the level of expression
 PT of a marker in a patient sample with that in the control non-breast
 PT cancer sample.
 XX
 PS Claim 1; SEQ ID NO 480; 128pp; English.
 XX
 CC The present invention describes a method for assessing whether a patient
 CC is afflicted with breast cancer. The method comprises comparing the level
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
 CC ABR47386 to ABR47632) in a patient sample and the normal level of
 CC expression of the marker in a control non-breast cancer sample, where a
 CC significant increase in the level of expression of the marker in the
 CC patient sample and the normal level is an indication that the patient is
 CC afflicted with breast cancer. The breast cancer associated sequences from
 CC the present invention have cytostatic activities and can be used in gene
 CC therapy. The method is useful for diagnosing and treating breast cancer.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences

SQ Sequence 206 AA;

Query Match 100.0%; Score 597; DB 6; Length 206;

Best Local Similarity 100.0%; Pred. No. 2.8e-60;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSGEPGQTSVAPPEEVEPGSGVRIWEYCEPCGFEATYLELASAVKEQYPGIEISRLG	60
DB	92	MSGEPGQTSVAPPEEVEPGSGVRIWEYCEPCGFEATYLELASAVKEQYPGIEISRLG	151
QY	61	GTGAFEIEINQOLVFSKLENGGFPYEKDLIAIRRASNGETLEKITNSRPPCVIL	115
DB	152	GTGAFEIEINQOLVFSKLENGGFPYEKDLIAIRRASNGETLEKITNSRPPCVIL	206

Search completed: March 27, 2006, 12:00:16

Job time : 121.82 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2006, 12:27:44 ; Search time 99.1617 Seconds

(without alignments)
484.566 Million cell updates/sec

Title: US-09-824-787B-2

Perfect score: 597

Sequence: 1 MSSEPGQTSVAPPEEVEPG.....ASNGETLEKITSRPPCVIL 115

Scoring table: BLOSUM62

Searched: 1867569 seqs, 41829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	115	3	US-09-824-787B-2
2	597	100.0	115	4	US-10-435-696-93
3	597	100.0	115	4	US-10-457-829-2
4	597	100.0	115	5	US-10-887-230-2
5	597	100.0	115	5	US-10-855-588-86
6	597	100.0	115	6	US-11-003-819-2
7	597	100.0	115	6	US-09-833-203-34
8	597	100.0	115	4	US-10-264-049-4187
9	597	100.0	115	4	US-09-925-301-966
10	597	100.0	115	4	US-10-457-829-155
11	597	100.0	115	4	US-10-177-293-480
12	403	67.5	90	4	US-10-457-829-153
13	148.5	24.9	95	6	US-11-097-143-38691
14	129	21.6	250	4	US-10-424-599-153780
15	126.5	21.2	244	4	US-10-739-930-10642
16	121	20.3	237	4	US-10-767-701-41097
17	119	19.9	232	4	US-10-437-963-106912
18	114	19.1	257	4	US-10-437-963-181063
19	113	18.9	250	4	US-10-767-701-44880
20	109	18.3	250	4	US-10-425-115-229236
21	108.5	18.2	228	4	US-10-424-599-149832
22	108.5	18.2	257	4	US-10-425-115-212160
23	87	14.6	198	6	US-11-097-143-7521
24	85.5	14.3	74	4	US-10-264-049-2688
25	80	13.4	137	3	US-09-374-046A-96
26	80	13.4	137	4	US-10-616-263-96
27	80	13.4	146	3	US-09-892-877-301

28	80	13.4	146	3	US-09-948-783-314	Sequence 314, App
29	80	13.4	163	4	US-10-614-853-14	Sequence 14, App1
30	79	13.2	469	4	US-10-369-493-22836	Sequence 22836, A
31	79	13.2	2042	4	US-10-437-963-112456	Sequence 112456, A
32	77	12.7	370	4	US-10-359-493-123317	Sequence 123317, A
33	76	12.7	282	4	US-10-437-963-136117	Sequence 136117, A
34	76	12.7	1463	4	US-10-437-963-163410	Sequence 163410, A
35	75.5	12.6	393	4	US-10-369-493-1082	Sequence 1082, Ap
36	75.5	12.6	590	4	US-10-437-963-114110	Sequence 114110, A
37	75	12.6	885	4	US-10-437-963-163413	Sequence 163413, A
38	74.5	12.5	198	4	US-10-424-599-172841	Sequence 172841, A
39	74.5	12.5	670	4	US-10-437-963-145442	Sequence 145442, A
40	74	12.4	394	4	US-10-425-114-11011	Sequence 11011, A
41	73.5	12.3	498	4	US-10-424-599-275620	Sequence 275620, A
42	73	12.2	531	4	US-10-437-963-185210	Sequence 185210, A
43	73	12.2	1303	5	US-10-450-763-42589	Sequence 42589, A
44	72.5	12.1	393	5	US-10-739-930-9981	Sequence 9981, Ap
45	72.5	12.1	795	4	US-10-334-143-19	Sequence 19, App1

ALIGNMENTS

```

RESULT 1
US-09-824-787B-2
Sequence 2, Application US/09824787B
Patent No. US20020155447A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
APPLICANT: Evans, Elizabeth E.
APPLICANT: Bottello, Melinda A.
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
FILE REFERENCE: 1821.0040001
CURRENT APPLICATION NUMBER: US/09/824, 787B
CURRENT FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/194,463
PRIOR FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 147
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 115
TYPE: PRT
ORGANISM: Homo sapiens
US-09-824-787B-2

Query Match      100.0%; Score 597; DB 3; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MSSEPGQTSVAPPEEVEPGGVIVVEYCEPGFATYELASVAYEYPGIEISRLG 60
      |||
DB      1 MSSEPGQTSVAPPEEVEPGGVIVVEYCEPGFATYELASVAYEYPGIEISRLG 60
      |||

QY      61 GTGAFFIEINGQLVFSKLENGGFPEYKDLIAIRRASNGETLEKITSRPPCVIL 115
      |||
DB      61 GTGAFFIEINGQLVFSKLENGGFPEYKDLIAIRRASNGETLEKITSRPPCVIL 115
      |||

RESULT 2
US-10-435-696-93
Sequence 93, Application US/10435696
Patent No. US20040018525A1
GENERAL INFORMATION:
APPLICANT: Wirtz, Ralph
APPLICANT: Munnes, Marc
APPLICANT: Kalibadis, Harald
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE PREDICTION, DIAGNOSIS, PROGNOSIS,
FILE REFERENCE: Lea 36 108
CURRENT APPLICATION NUMBER: US/10/435, 696
CURRENT FILING DATE: 2003-05-09
PRIOR APPLICATION NUMBER: EP03003112.4

```

PRIOR FILING DATE: 2003-02-13
PRIOR APPLICATION NUMBER: EP02010291.9
PRIOR FILING DATE: 2002-05-21
NUMBER OF SEQ ID NOS: 314
SOFTWARE: PatentIn version 3.1
SEQ ID NO 93
LENGTH: 115
TYPE: PRT
ORGANISM: Homo sapiens
US-10-435-696-93

Query Match 100.0%; Score 597; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
DB 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
OY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115

RESULT 3
US-10-457-829-2
Sequence 2, Application US/10457829
Publication No. US20040063907A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
APPLICANT: Evans, Elizabeth E.
APPLICANT: Borrello, Melinda A.
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
FILE REFERENCES: 1821.0040005
CURRENT APPLICATION NUMBER: US/10/457,829
CURRENT FILING DATE: 2003-06-10
PRIOR APPLICATION NUMBER: US 60/464,650
PRIOR FILING DATE: 2003-04-23
NUMBER OF SEQ ID NOS: 160
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 115
TYPE: PRT
ORGANISM: Homo sapiens
US-10-457-829-2

Query Match 100.0%; Score 597; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
DB 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
OY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115

RESULT 4
US-10-887-230-2
Sequence 2, Application US/10887230
Publication No. US20050042218A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
TITLE OF INVENTION: HGC Class I - Peptide-Antibody Conjugates with Modified
FILE REFERENCES: 1843.0160002
CURRENT APPLICATION NUMBER: US/10/887,230
CURRENT FILING DATE: 2004-07-09
PRIOR APPLICATION NUMBER: US 60/485,716
PRIOR FILING DATE: 2003-7-10

PRIOR APPLICATION NUMBER: US 60/513,043
PRIOR FILING DATE: 2003-10-22
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2
LENGTH: 115
TYPE: PRT
ORGANISM: Homo sapiens
US-10-887-230-2

Query Match 100.0%; Score 597; DB 5; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
DB 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
OY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115

RESULT 5
US-10-855-588-86
Sequence 86, Application US/10855588
Publication No. US20050042642A1
GENERAL INFORMATION:
APPLICANT: Monahan, John
APPLICANT: Hoersch, Sebastian
APPLICANT: Anderson, Dustin
APPLICANT: Endege, Wilson
APPLICANT: Ford, Donna
APPLICANT: Glatt, Karen
APPLICANT: Gordatcheva, Bella
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Xu, Yong Yao
APPLICANT: Ganavarapu, Manjula
APPLICANT: Zhao, Xumei
APPLICANT: Robert Schlegel
APPLICANT: Maureen Mertens
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
BREAST CANCER
FILE REFERENCES: MRI-064
CURRENT APPLICATION NUMBER: US/10/855,588
CURRENT FILING DATE: 2004-05-26
PRIOR APPLICATION NUMBER: 60/474,281
PRIOR FILING DATE: 2003-05-29
PRIOR APPLICATION NUMBER: 60/555,557
PRIOR FILING DATE: 2004-03-24
NUMBER OF SEQ ID NOS: 96
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 86
LENGTH: 115
TYPE: PRT
ORGANISM: human
US-10-855-588-86

Query Match 100.0%; Score 597; DB 5; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
DB 1 MSGPGQTSVAPPEEVEPSSGVRIIVEYCEPCGFATYELASAVKEQYPGIEIESRLG 60
OY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115

RESULT 6

US-11-003-819-2
; Sequence 2, Application US/11003819
; Publication No. US20050158323A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Elizabeth E.
; APPLICANT: Paris, Mark U.
; APPLICANT: Sahasrabudhe, Deepak M.
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Methods of Killing Tumor Cells by Targeting Internal Antigens
; FILE REFERENCE: 1843.0190002
; CURRENT FILING DATE: US/11/003,819
; PRIOR FILING DATE: 2004-12-06
; PRIOR APPLICATION NUMBER: US 60/256,572
; PRIOR FILING DATE: 2003-12-04
; PRIOR APPLICATION NUMBER: US 60/531,688
; PRIOR FILING DATE: 2003-12-23
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-003-819-2

Query Match 100.0%; Score 597; DB 6; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 60
DB 1 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 60
QY 61 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115

RESULT 7
US-09-833-203-34
; Sequence 34, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT FILING DATE: US/09/833,203
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35
US-09-833-203-34

Query Match 100.0%; Score 597; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 2.1e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 60
DB 3 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 62
QY 61 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115
DB 61 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115

DB 63 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 117
RESULT 8
US-10-264-049-4187
; Sequence 4187, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA3391
; CURRENT FILING DATE: US/10/264,049
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 4187
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-049-4187

Query Match 100.0%; Score 597; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.3e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 60
DB 10 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 69
QY 61 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115
DB 70 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 124

RESULT 9
US-09-925-301-966
; Sequence 966, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT FILING DATE: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 966
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-966

Query Match 100.0%; Score 597; DB 3; Length 131;
Best Local Similarity 100.0%; Pred. No. 2.4e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 60
DB 17 MSBPGQTSVAPPEVEPGSGVRIVVEYCEPCGFATYLELASAVKQYPGIEISRLG 76
QY 61 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115
DB 77 GTGAFEIINGQLVFSKLENGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 131

RESULT 10
US-10-457-829-155
Sequence 155, Application US/10457829
Publication No. US20040063907A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
APPLICANT: Evans, Elizabeth E.
APPLICANT: Borrello, Melinda A.
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
FILE REFERENCE: 1821.0040005
CURRENT APPLICATION NUMBER: US/10/457, 829
CURRENT FILING DATE: 2003-06-10
PRIOR APPLICATION NUMBER: US 60/464,650
PRIOR FILING DATE: 2003-04-23
NUMBER OF SEQ ID NOS: 160
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 155
LENGTH: 131
TYPE: PRT
ORGANISM: Homo sapiens
US-10-457-829-155

Query Match 100.0%; Score 597; DB 4; Length 131;
Best Local Similarity 100.0%; Pred. No. 2.4e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGERGQTSVAPPEPEVPGSGVRIVYECPCGFEATYLELASAVKQYPGIEISRLG 60
DB 17 MSGERGQTSVAPPEPEVPGSGVRIVYECPCGFEATYLELASAVKQYPGIEISRLG 76
QY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115
DB 77 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 131

RESULT 11
US-10-177-293-480
Sequence 480, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Zhao, Karen
APPLICANT: Gao, Xumei
APPLICANT: Gannavaru, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Naureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Baat Jr., Robert C.
APPLICANT: Horobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177, 293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299, 887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301, 572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306, 501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325, 002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362, 585

PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 480
LENGTH: 206
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-480

Query Match 100.0%; Score 597; DB 4; Length 206;
Best Local Similarity 100.0%; Pred. No. 4.4e-58;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGERGQTSVAPPEPEVPGSGVRIVYECPCGFEATYLELASAVKQYPGIEISRLG 60
DB 92 MSGERGQTSVAPPEPEVPGSGVRIVYECPCGFEATYLELASAVKQYPGIEISRLG 151
QY 61 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 115
DB 152 GTGAFEIINGQLVFSKLENGGFPYEKDLIEAIRASNGETLEKITSRPPCVIL 206

RESULT 12
US-10-457-829-153
Sequence 153, Application US/10457829
Publication No. US20040063907A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
APPLICANT: Evans, Elizabeth E.
APPLICANT: Borrello, Melinda A.
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
FILE REFERENCE: 1821.0040005
CURRENT APPLICATION NUMBER: US/10/457, 829
CURRENT FILING DATE: 2003-06-10
PRIOR APPLICATION NUMBER: US 60/464,650
PRIOR FILING DATE: 2003-04-23
NUMBER OF SEQ ID NOS: 160
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 153
LENGTH: 90
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE: MISC_FEATURE
NAME/KEY: MISC_FEATURE
LOCATION: (18)..(18)
OTHER INFORMATION: Xaa is an unknown amino acid
FEATURE: MISC_FEATURE
NAME/KEY: MISC_FEATURE
LOCATION: (46)..(46)
OTHER INFORMATION: Xaa is an unknown amino acid
FEATURE: MISC_FEATURE
NAME/KEY: MISC_FEATURE
LOCATION: (78)..(78)
OTHER INFORMATION: Xaa is an unknown amino acid
US-10-457-829-153

Query Match 67.5%; Score 403; DB 4; Length 90;
Best Local Similarity 87.6%; Pred. No. 6.5e-37;
Matches 78; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSGERGQTSVAPPEPEVPGSGVRIVYECPCGFEATYLELASAVKQYPGIEISRLG 60
DB 1 MSGERGQTSVAPPEPEVPGSGVRIVYECPCGFEATYLELASAVKQYPGIEISRLG 60
QY 61 GTGAFEIINGQLVFSKLENGGFPYEKDL 89
DB 61 GTGAFEIINGQLVFSKLENGGFPYEKDV 89

RESULT 13

US-11-097-143-38691
; Sequence 38691, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: C1000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38691
; LENGTH: 95
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-38691

Query Match 24.9%; Score 148.5; DB 6; Length 95;
Best Local Similarity 34.0%; Pred. No. 1.8e-08;
Matches 32; Conservative 17; Mismatches 44; Indels 1; Gaps 1;

QY 23 VIVVEYCEPCGFEATYELASAVEQYPGIIESRLGCTGAFELINQGVFSKLENGG 82
DB 2 VVVEVEYCGICNFCSCCHLREFLLASPDLDISCRTRGSGFEVSIQGLVHSLSCIA 61
QY 83 FPEKDLLEAIRRANGTLEKITNSR-PPCVIL 115
DB 62 FPGHASYLAQVOKARGSPVEKVLDPKPCVVM 95

RESULT 14
US-10-424-599-153780
; Sequence 153780, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 153780
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109887C.1.pep
US-10-424-599-153780

Query Match 21.6%; Score 129; DB 4; Length 250;

Best Local Similarity 22.0%; Pred. No. 9.7e-06;
Matches 38; Conservative 20; Mismatches 27; Indels 88; Gaps 4;

QY 13 PPEE-----VPGSGVRIVVEYCEPCGFEATYELASAVEQYPGIEI----- 55
DB 78 PPEKTNINATPPGLANTVHINFCSSCYKGTAVTKMMLAIFPCTEVILANYPPTLPKR 137
QY 56 -----ESR 58
DB 138 LLSKLVVVOIGVIGVAVVASHIFPMLGFVAPPPYVNLRRNFCTIASTMLGNALQSF 197
QY 59 LGGTGAFPIEINGQVFSKLENGGFPYE--KDLLEAIRRANGTLEKITNS 108
DB 198 LQSSGAFEVYENGELVFSKLEKGRFPGEIEIKDLI-----TKMTNS 239

RESULT 15
US-10-739-930-10642
; Sequence 10642, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11068
; SEQ ID NO 10642
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Trifolium aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-CS5262_1.p
US-10-739-930-10642

Query Match 21.2%; Score 126.5; DB 5; Length 244;
Best Local Similarity 21.7%; Pred. No. 1.8e-05;
Matches 40; Conservative 21; Mismatches 36; Indels 87; Gaps 5;

QY 4 EPGQTSVAPPEEVE-----PSSGVRIVVEYCEPCGFEATYELASAVEQYPG 52
DB 46 QPQDSS---PDADDSQALAEQVDGPGSGTVELKFCASCYSKGNAMTKRMLDTSPPG 101
QY 53 IEI----- 55
DB 102 IHVLENYPPFPFKRALGKVPFVQVCAIATLMAQDQIFPFRGMVPPMYTISLRANREGT 161
QY 56 -----ESRLGCTGAFPIEINGQVFSKLENGGFPYEKDLLEAI-RRASN---GE 100
DB 162 MASVMMFNGFASQSLQSSGAFEVYENGELVFSKLEKGRFPGEIEIKDLI----- 221
QY 101 TLEK 104
DB 222 NLEK 225

Search completed: March 27, 2006, 12:37:11
Job time : 100.162 secs